

IEEE Standard Definitions and Characterization of Floating Gate Semiconductor Arrays

Sponsor

**Standards Committee
of the
IEEE Electron Devices Society**

Approved 25 June 1998

IEEE-SA Standards Board

Abstract: This standard describes the underlying physics and the operation of floating gate memory arrays, specifically, UV erasable EPROM, byte rewritable E²PROMs, and block rewritable “flash” EEPROMs. In addition, reliability hazards are covered with focus on retention, endurance, and disturb. There are also clauses on the issues of testing floating gate arrays and their hardness to ionizing radiation.

Keywords: disturb, endurance, E²PROM, EEPROM, EPROM, floating gate memories, nonvolatile memory, radiation hardness, reliability, retention, testing

The Institute of Electrical and Electronics Engineers, Inc.
345 East 47th Street, New York, NY 10017-2394, USA

Copyright © 1999 by the Institute of Electrical and Electronics Engineers, Inc.
All rights reserved. Published 9 February 1999. Printed in the United States of America.

Print: ISBN 0-7381-0333-0 SH94655
PDF: ISBN 0-7381-0620-8 SS94655

No part of this publication may be reproduced in any form, in an electronic retrieval system or otherwise, without the prior written permission of the publisher.

IEEE Standards documents are developed within the IEEE Societies and the Standards Coordinating Committees of the IEEE Standards Association (IEEE-SA) Standards Board. Members of the committees serve voluntarily and without compensation. They are not necessarily members of the Institute. The standards developed within IEEE represent a consensus of the broad expertise on the subject within the Institute as well as those activities outside of IEEE that have expressed an interest in participating in the development of the standard.

Use of an IEEE Standard is wholly voluntary. The existence of an IEEE Standard does not imply that there are no other ways to produce, test, measure, purchase, market, or provide other goods and services related to the scope of the IEEE Standard. Furthermore, the viewpoint expressed at the time a standard is approved and issued is subject to change brought about through developments in the state of the art and comments received from users of the standard. Every IEEE Standard is subjected to review at least every five years for revision or reaffirmation. When a document is more than five years old and has not been reaffirmed, it is reasonable to conclude that its contents, although still of some value, do not wholly reflect the present state of the art. Users are cautioned to check to determine that they have the latest edition of any IEEE Standard.

Comments for revision of IEEE Standards are welcome from any interested party, regardless of membership affiliation with IEEE. Suggestions for changes in documents should be in the form of a proposed change of text, together with appropriate supporting comments.

Interpretations: Occasionally questions may arise regarding the meaning of portions of standards as they relate to specific applications. When the need for interpretations is brought to the attention of IEEE, the Institute will initiate action to prepare appropriate responses. Since IEEE Standards represent a consensus of all concerned interests, it is important to ensure that any interpretation has also received the concurrence of a balance of interests. For this reason, IEEE and the members of its societies and Standards Coordinating Committees are not able to provide an instant response to interpretation requests except in those cases where the matter has previously received formal consideration.

Comments on standards and requests for interpretations should be addressed to:

Secretary, IEEE-SA Standards Board
445 Hoes Lane
P.O. Box 1331
Piscataway, NJ 08855-1331
USA

Note: Attention is called to the possibility that implementation of this standard may require use of subject matter covered by patent rights. By publication of this standard, no position is taken with respect to the existence or validity of any patent rights in connection therewith. The IEEE shall not be responsible for identifying patents for which a license may be required by an IEEE standard or for conducting inquiries into the legal validity or scope of those patents that are brought to its attention.

Authorization to photocopy portions of any individual standard for internal or personal use is granted by the Institute of Electrical and Electronics Engineers, Inc., provided that the appropriate fee is paid to Copyright Clearance Center. To arrange for payment of licensing fee, please contact Copyright Clearance Center, Customer Service, 222 Rosewood Drive, Danvers, MA 01923 USA; (978) 750-8400. Permission to photocopy portions of any individual standard for educational classroom use can also be obtained through the Copyright Clearance Center.

Introduction

(This introduction is not part of IEEE Std 1005-1998, IEEE Standard Definitions and Characterization of Floating Gate Semiconductor Arrays.)

Over the last few years, floating gate memory arrays have become the overwhelming technology of choice for those applications that require nonvolatile semiconductor memory. Because this technology raises issues heretofore not commonly addressed by users, however, there has been uncertainty as to how to best characterize and evaluate products that incorporate it for various applications. A complicating factor has been the fact that products that use this technology have been available from a number of commercial suppliers who have employed several variants of the basic technology and who have sometimes used inconsistent terminology to describe it.

This standard is intended for users or prospective users of floating gate technology as well as for those who are new to the development or manufacturing of it. This standard provides an introduction to the technology, describes the variants in current use, addresses the salient issues, and describes tests that are designed to characterize and evaluate products in light of these issues. It also defines common terms in order to facilitate communication.

This standard is a revision of IEEE Std 1005-1991. IEEE Std 1005-1991 was restricted in coverage to byte rewritable E²PROMs. The initial edition was issued with the knowledge that block rewritable, or “flash,” nonvolatile memories were becoming available. The intention at the time of initial issuance was to issue a revision that would include flash memories. With this revision, that intention has been carried to fruition. In addition, this revision has been broadened to include UV-EPROMs.

There is very extensive literature on floating gate nonvolatile devices that describes a tremendous variety of different floating gate cells that can be used to perform memory functions. This standard does not attempt to be encyclopedic. Rather, descriptions in this standard are limited to cells that are in current manufacturing production at the time the standard was written.

The members of the working group would like to acknowledge to contribution of Richard Wiker, who served as the first chair person of the working group, but was forced to drop out when his employer drastically reduced its support for professional activities.

At the time this standard was completed, the Working Group on Floating Gate Arrays had the following membership:

John Caywood, *Chair*

Edward Dollar
Michael Knoll
Herman Maes

Walt Rau
John Schreck
David Sweetman

Michael Van Buskirk
Richard Wegener
Kuniyoshi Yoshikawa

The following persons were on the balloting committee:

Radu M. Barsan
William D. Brown
John M. Caywood
Michael. Chen
Christophe Chevallier
Roy A. Colclaser
Sameer S. Haddad
Fred B. Jenne
Dah-Bin Kao
Sohrah Kianian
Michael Knoll
Al V. Kordesch
Vikram Kowshik

Di-Son Kuo
Wlodek Kurjanowicz
Al Kwok
Bob Lee
Wei-Hua Liu
Steven W. Longcor
Tso-Ping Ma
Peter N. Manos
Marco Mastrapasqua
Sharon K. Murray
William H. Owen
Wayne Paulson

Paolo Pavan
Frank Qian
Tamer San
James D. Sansbury
Robert Schneider
Ritu Shrivastava
Yasushi Terada
Tom Toombs
Hai Wang
Ping Wang
H.A. Richard Wegener
Kuniyoshi Yoshikawa
Jain-Chen Young

When the IEEE-SA Standards Board approved this standard on 25 June 1998, it had the following membership:

Richard J. Holleman, *Chair*

Donald N. Heirman, *Vice Chair*

Judith Gorman, *Secretary*

Satish K. Aggarwal
Clyde R. Camp
James T. Carlo
Gary R. Engmann
Harold E. Epstein
Jay Forster*
Thomas F. Garrity
Ruben D. Garzon

James H. Gurney
Jim D. Isaak
Lowell G. Johnson
Robert Kennelly
E. G. "Al" Kiener
Joseph L. Koepfinger*
Stephen R. Lambert
Jim Logothetis
Donald C. Loughry

L. Bruce McClung
Louis-François Pau
Ronald C. Petersen
Gerald H. Peterson
John B. Posey
Gary S. Robinson
Hans E. Weinrich
Donald W. Zipse

*Member Emeritus

Contents

1.	Overview.....	1
1.1	Scope.....	1
1.2	Document organization.....	1
1.3	Basic physics of floating gate nonvolatile devices	2
1.4	Introduction to floating gate memory arrays	17
2.	References.....	18
3.	Definitions, abbreviations and acronyms, and symbols.....	19
3.1	Definitions.....	19
3.2	Acronyms and abbreviations.....	26
3.3	Symbols	27
4.	Basic floating gate FET types	27
4.1	Background.....	27
4.2	EPROM.....	28
4.3	E ² PROM	40
4.4	Flash memory cells	44
5.	Reliability considerations.....	75
5.1	Background.....	75
5.2	Failure rate components unique to floating gate devices.....	75
5.3	MOS failure mechanisms.....	76
5.4	Effect of retention on failure rate.....	76
5.5	Effect of endurance on failure rate.....	78
5.6	Determination of composite failure rate	79
5.7	Calculation of the composite failure rate	79
5.8	Conclusion	81
6.	Retention.....	81
6.1	Background.....	81
6.2	Concept of retention.....	82
6.3	Verification of retention.....	84
6.4	Discharge of capacitor by F-N tunneling.....	85
7.	Endurance	86
7.1	Background.....	86
7.2	Endurance concept.....	86
7.3	Endurance rating objectives.....	89
7.4	Endurance verification approaches.....	90
8.	Disturbs.....	92
8.1	Introduction.....	92
8.2	Basic disturb mechanisms.....	92

8.3	Program and erase disturbs	96
8.4	Effects enhancing disturbs	99
9.	Testing methodology	100
9.1	Background	100
9.2	Failure definition	100
9.3	Methods of tests	101
9.4	Write conditions	101
9.5	Power-supply sequence of EEPROMs	103
9.6	Test patterns	103
9.7	Retention test conditions	104
9.8	Endurance test conditions	105
9.9	Appendix	106
10.	Ionizing radiation effects on floating gate memory ICs	108
10.1	Radiation effects background	108
10.2	Measurement of radiation effects	110
10.3	Total dose tests	111
10.4	Dose-rate tests	112
10.5	SEP tests	114
11.	Other nonvolatile semiconductor structures	115
11.1	Background	115
11.2	Other semiconductor nonvolatile technologies	115
11.3	Non-memory applications of nonvolatile floating gate technology	116
12.	Bibliography	117

IEEE Standard Definitions and Characterization of Floating Gate Semiconductor Arrays

1. Overview

1.1 Scope

This standard is restricted to the consideration of floating gate memory arrays. Metal-nitride-oxide semiconductors (MNOS) memory arrays are already covered in IEEE Std 641-1987.

This standard does not deal with the use of any floating gate memory cell in non-memory applications such as logic or analog functions.

Such excluded topics are mentioned in Clause 11.

1.2 Document organization

This standard has eleven clauses.

Clause 1 contains an introduction to the physics unique to floating gate memory arrays and an overview of typical array architectures. Clause 1 also defines the scope of this standard.

Clause 2 provides references that contain added detail on specific concepts.

Clause 3 defines the definitions, abbreviations, acronyms, and symbols that are used in the remaining clauses.

Clause 4 describes the variations on the basic floating gate nonvolatile transistor structure that have been used in commercially available devices. This revision has been expanded to include both ultraviolet-erasable programmable read-only memories (UV-EPROMs) and the “flash” electrically erasable programmable read-only memory (EEPROM) devices in addition to the byte-alterable electrically erasable programmable read-only memory (E²PROM) arrays that were covered in the first release of this standard.

Clause 5 explores the various reliability considerations that are involved with these devices. It discusses retention and endurance failures, and the interaction between endurance, retention, and standard semiconductor failure mechanisms in determining the device failure rate.

Clause 6 defines the retention of data stored in these arrays. It describes how to specify and perform engineering verification of retention. Some typical production screens and monitors are indicated.

Clause 7 defines the endurance of floating gate arrays. Effects which limit the endurance are discussed. The specification and engineering verification of endurance are described.

Clause 8 is devoted to the mechanisms that can “disturb” the charge on a floating gate in the course of the operation of the array. This is a subject that has gained importance with the rise to prominence of flash EEPROM arrays.

Clause 9 deals with the testing of floating gate arrays. It describes the more common features incorporated into these arrays and discusses how to efficiently test these complex products.

Clause 10 discusses the effects that various forms of ionizing radiation may have on floating gate arrays and the approaches that are used to test for these effects.

Clause 11 sketches briefly the use of floating gate cells in non-memory component applications. Competing approaches to electrically alterable, semiconductor non-volatile memory are also mentioned.

1.3 Basic physics of floating gate nonvolatile devices

The floating gate memory element consists of an electrode that is completely encased in an insulator. Because this electrode has no direct electrical connection to any other electrical conductor, it is often referred to as a floating gate. The electrode acts as the gate electrode to a metal-oxide semiconductor field-effect transistor (MOSFET) that serves as the sense transistor. The potential on the floating gate can be modulated by other adjacent conductors that capacitively couple to the floating gate. The potential on the floating gate can be sensed by its effect on the current flow through the MOSFET. This potential can be altered by changing the number of electrons on the gate. Altering the logical output by changing the charge state on the floating gate is referred to as “writing” to the device.

The fundamental challenge in designing a reprogrammable nonvolatile memory cell is the necessity of finding a controllable, reproducible effect that has a sufficiently nonlinear current-voltage characteristic so that the cell can be written to a state at one voltage in $<<1$ s and the state can be maintained during read or storage at another voltage for more than 10 years ($\sim 3 \times 10^8$ s). For example, an EPROM may be written in 10 μ s with 12 V on the control gate and 6 V on the drain and may be expected to retain data while being read with 5 V on the control gate and 1.2 V on the drain. Two physical phenomena are known to possess the necessary properties: hot carrier injection to an electrically isolated gate electrode and Fowler-Nordheim tunneling between an electrically isolated electrode and another conducting electrode. These are the phenomena that are used for writing in present day commercial UV-EPROMs, byte-alterable E²PROMs, and flash EEPROMs. These two phenomena will be discussed, in turn, in the clauses that follow. This description is limited to those phenomena that are specific to floating gate arrays. For general background physics, refer to a standard text on device or solid-state physics.

1.3.1 Hot carrier injection

As mentioned previously, the floating gate electrode is completely encased in an insulator. The wide-band gap of the insulator provides a potential barrier to prevent electrons from flowing onto or off of the floating gate, as is shown in Figure 1. Hot carrier injection occurs when carriers are accelerated to a high enough energy level to surmount the barrier. There are several mechanisms that may accelerate the carriers, but acceleration of electrons by lateral fields in the channel is the only hot electron mechanism that is currently used by floating gate memories for writing. Consequently, the discussion in this standard will be restricted to hot carrier injection from channel hot electrons (CHE).

In order for a lateral channel field to inject an electron onto the floating gate, several events must occur. First, the electron must be accelerated in the field to an energy level that is in excess of the barrier height. Next, the electron must be scattered so that it is directed toward the floating gate. The electron must reach the oxide interface without its energy dropping below the oxide barrier height and cross the Si/SiO₂ interface and surmount the oxide potential barrier (see Hu [B38] and Tam, et al. [B102]).

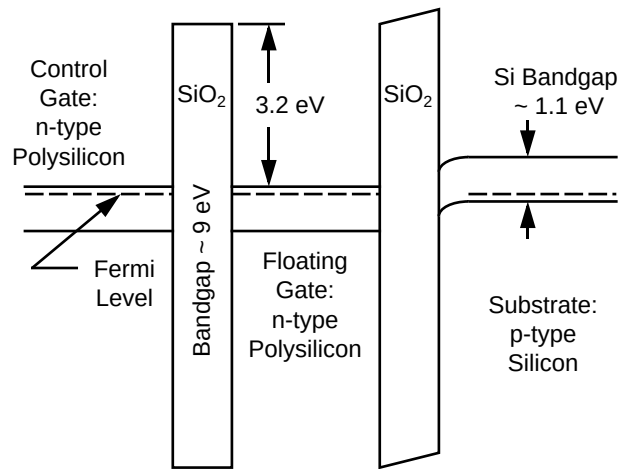


Figure 1—Band structure diagram of floating gate structure showing the n-type polysilicon control gate on the left, the n-type polysilicon floating gate in the center sandwiched between two SiO₂ layers, and the p-type substrate on the right. The high barrier between the conduction band in the silicon and the SiO₂ inhibits electrons from passing from the silicon into the SiO₂ in normal equilibrium conditions.

CHEs can be generated at either the drain or the source of the floating gate, depending on the electrode configuration. Injecting electrons onto the gate will be referred to as programming in this standard. Drain-side programming is more common than source-side programming and will be described first.

1.3.1.1 Drain-side programming

Drain-side programming is commonly employed because it can function with a very simple structure. All that is required is a gate with a drain and a source. This simplicity lends itself to a small cell size.

For programming to occur, the drain and floating gate are biased at relatively high voltages. The drain bias is applied with a direct connection to a voltage source, while the gate bias depends upon capacitive coupling. The potential that is induced on the floating gate by a change in one of the surrounding electrodes depends on the ratio of the capacitance between the floating gate and the electrode in question to the total floating gate capacitance. For this reason, the operation of a floating gate memory cell is often described in terms of coupling ratios. For example, the ratio of the capacitance between the floating gate and the control gate to the total floating gate capacitance is called the control gate coupling ratio.

For efficient programming, the transistor should be biased in saturation so that a high lateral electric field is created in the depletion region near the pinch-off point. In Figure 2(a), a gate is shown to be biased in such a state. The channel inversion layer is wider near the source and narrows as it approaches the pinch-off point. As the electrons pass through the pinch-off point, they are strongly accelerated in the high field of the drain-depletion region. Some of the hot electrons are scattered in a direction such as to reach the Si/SiO₂ interface (see Eitan and Frohman-Bentchkowsky[B22]).

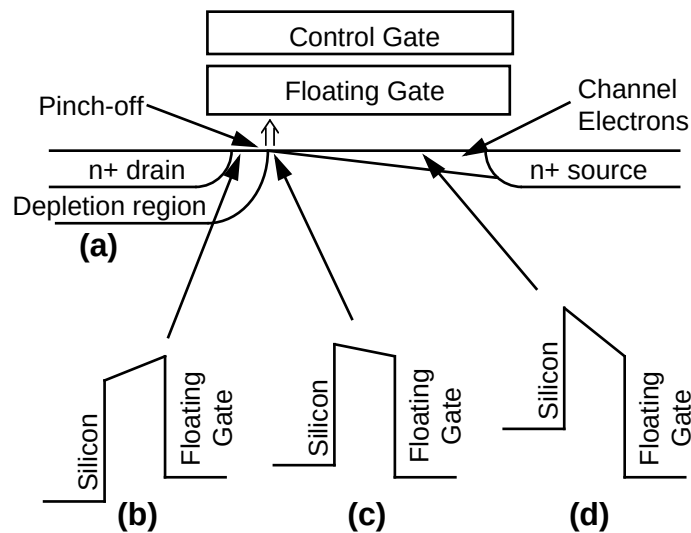


Figure 2—(a) Illustrates the physical layout and indicates the injection of electrons near the pinch-off point. (b)–(d) show the potential diagrams of the conduction bands at the points labeled (b)–(d) in (a). Note that, near the drain, the field in the oxide repels electrons injected over the barrier from the substrate.

The shape of the SiO_2 barrier varies along the channel length because the potential along the channel varies from source to drain while the gate potential remains constant. Figures 2(b), 2(c), and 2(d) show schematically the oxide barrier shape at three selected points along the channel. Figure 2(b) shows that the electric field in the gate oxide near the drain is repulsive, which greatly reduces the collection probability at this point. Near the pinch-off point, the oxide field tends to collect electrons, as shown in Figure 2(c). Near the source, the oxide field is very strongly biased toward collection, as shown in Figure 2(d).

The highest density of hot electrons is toward the drain, where the oxide field is repulsive. Near the source, where the oxide field is very strongly collective, there are almost no hot electrons. Only near the pinch-off point can hot electrons and a collecting oxide field be found in conjunction. As the floating gate is programmed, the potential on the floating gate becomes lower in magnitude, and the pinch-off point (and consequently the point of maximum electron injection) moves from near the drain towards the source. Because only a small portion of the channel length is effective for programming, the programming efficiency is rather low. The portion of the hot carriers created that are collected on the floating gate is typically in the range of 10^{-5} to 10^{-6} for drain-side CHE programming.

It should be noted, however, that as metal-oxide semiconductor (MOS) technology scales down, the drain current becomes a better generator of hot electrons. Additionally, because the floating gate capacitance decreases with scaling, fewer electrons are required to cause a given amount of threshold shift. The result of these effects is that cells that program in times of 1–10 μs with drain currents in the range of 0.5–1 mA are presently common.

The programming of floating gate cells by drain-side injection of hot electrons is discussed in more detail in Fiegna, et al. [B26] and Tanaka and Ishikawa [B103].

1.3.1.2 Source-side injection

While the drain-side injection approach sacrifices electron injection efficiency in the interest of a simple cell, the source-side injection approach achieves a higher injection efficiency by separating the hot electron generation and collection functions. The price paid for the improved efficiency is a cell with more elements. This approach demands two gates that are adjacent to the gate oxide with two channels.

Figure 3 schematically illustrates this approach. There are two gates that act as transistor gates and that are referred to as the floating gate and the injection gate in Figure 3. These two gates are separated from each other by an insulating spacer (usually 50–100 nm thick). A direct electrical connection is made to the injection gate so that its potential is set directly. The potential on the floating gate is manipulated, as always, with capacitive coupling from the control gate. For programming, the drain and the floating gate are biased at relatively high potentials, while the injection gate is biased only slightly above threshold. The consequence of these biases is that the whole channel length under the floating gate is depleted, which forms a virtual drain on the drain side of the injection gate. A large lateral field is formed across the insulator that separates the two gates. The result is that the electrons emerge into the depletion region with high energy.

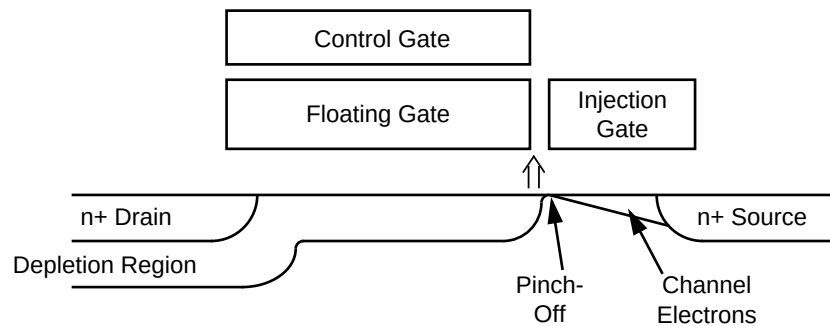


Figure 3—Schematic illustration of a generalized source-side injection cell that shows the pinch-off of the channel at the end of the injection gate and the depletion region that extends under the length of the floating gate to form a virtual drain at the drain end of the injection gate. The broad arrow indicates where the electrons are injected at the source end of the floating gate.

Because the injection gate device has a virtual drain, it can be made with a significantly shorter channel than a similar scale transistor that has a metallurgical drain junction without encountering punch-through. This allows the injection transistor to operate with a much lower voltage drop along the channel than is possible for drain-side programming. A consequence of all of this is that the source-side injection cell can operate with a lower drain voltage and still obtain the same electron energy distribution that the drain-side cell obtains at a higher drain voltage. The lower drain voltage leads to more favorable electron collection conditions at the floating gate Si/SiO₂ interface. Because of this, the injection efficiencies are two to three orders of magnitude higher than those for the drain-side injection cells.

For more details on source-side injection physics, see Kobayashi, et al. [B56], van Houdt, et al. [B111], and Wu, et al. [B118].

1.3.1.3 Split-gate cell

Split-gate cell is a term that is given to a cell in which the channel of the memory transistors is overlaid by two distinct gates, so that the transistor may be viewed as two transistors with channel lengths l_1 and l_2 that are connected through a virtual source/drain region. Such a cell is shown in the cross-section in Figure 4.

For all EPROM cells, the floating gate potential depends on the gate charge and the voltages coupled to it from the surrounding electrodes, which are usually the drain, control gate, and the body (or substrate). In the case of a cell with a single transistor, such as the one illustrated in Figure 2, drain coupling is generally detrimental and is minimized. Drain coupling is detrimental for a single gate cell because drain voltage coupled from the bit-lines to the cells on the unselected rows tends to bias these cells into conduction and enhance punch-through. This effect increases the minimum allowed effective channel length, which, in turn, decreases the electron injection efficiency when programming.

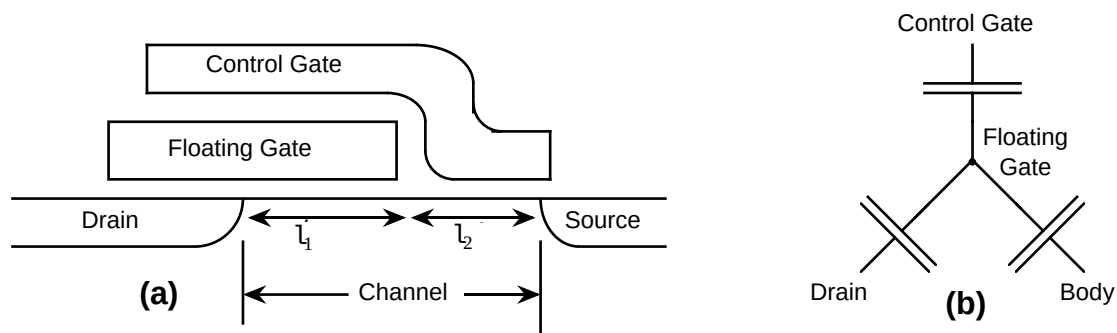


Figure 4—(a) Cross-section through a split gate cell showing the two portions of the channel, l_1 and l_2 ; and (b) schematic representation of the coupling capacitors that “steer” the floating gate

In the case of the split-gate cell, the grounded control gate inhibits punch-through on the unselected rows so that the drain overlap is much less significant. Some split-gate cells deliberately enhance the drain overlap and use the drain coupling to increase the floating gate potential for programming. The split-gate cell may be, and has been, applied to cells that employ both drain-side and source-side programming.

1.3.2 Fowler-Nordheim (F-N) tunneling

Another phenomenon that exhibits the necessary nonlinearity for writing is F-N tunneling, which was first described by Fowler and Nordheim in 1928 [B27] and was subsequently named after them. This effect can be used for either programming or erasing a floating gate. Some cells employ F-N tunneling for both programming and erasing, while others depend on CHE injection for programming and F-N tunneling for erasure only.

The basic idea is illustrated in Figure 5. The energy difference between the conduction and valence bands in Si is about 1.1 eV. The energy difference between these bands in SiO_2 is approximately 9 eV. When the two materials are joined, the conduction band in SiO_2 is 3.25 eV above that in Si. The differences in valence band energies is even larger (>4 eV). Since the thermal energy of an electron averages only 0.025 eV at room temperature, the chance of an electron in silicon gaining enough thermal energy to surmount the barrier and enter the conduction band in SiO_2 is exceedingly small. This case is illustrated in Figure 5(a).

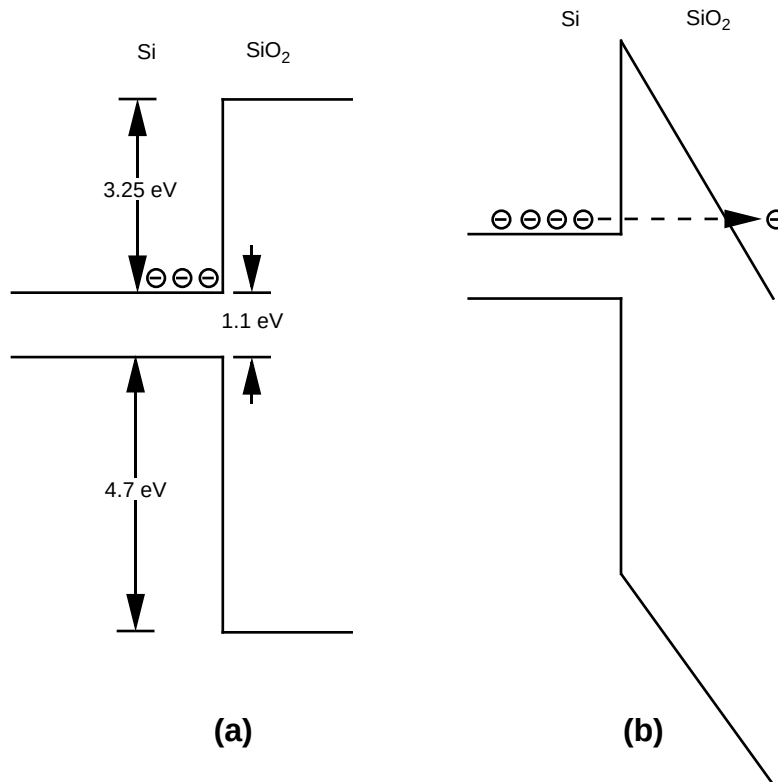


Figure 5—Potential diagram of the Si/SiO₂ interface (a) with no electric field and (b) with a strong applied electric field. The electric field applied in (b) is strong enough to allow electrons to tunnel from the conduction band in the Si into the conduction band in the SiO₂.

Fowler and Nordheim noted that, in the presence of a high electric field, the energy bands will be distorted, as illustrated in Figure 5(b). Under these conditions, there is a small but finite probability that an electron in the conduction band in the silicon will quantum-mechanically tunnel through the energy barrier and emerge in the conduction band of the SiO₂, as illustrated in Figure 5(b).

F-N emission was observed early in this century for the case of electrons being emitted from metals into a vacuum. In 1969, Lenzlinger and Snow [B66] observed this phenomenon for the Si-SiO₂ system. The F-N current increases exponentially with applied field, and becomes readily observable (i.e., $J \sim 10^{-6}$ A/cm²) for the Si-SiO₂ system when the field at the Si-SiO₂ interface is on the order of 10 MV/cm.

Local fields of this magnitude at voltages that are practicable for use on high-density integrated circuits (IC) have been obtained either by applying a voltage across a very thin (~10 nm) oxide layer grown on a smooth Si substrate or by applying a voltage across a somewhat thicker oxide (~30–100 nm) grown on polycrystalline electrodes. In the latter case, the field enhancement that arises from the positive curvature regions of the textured polycrystalline surface results in tunneling at similar voltages as in the first case (see Ellis, et al. [B23]). The case in which the field is applied across a thin oxide grown on the substrate can be further divided into the case in which the field is applied uniformly across the whole thin oxide region, and the case in which the field is applied to only a portion of the thin oxide region.

Figure 6 illustrates the potential and electric field lines for the geometrically enhanced tunneling case noted above. There are other structures that will result in this effect; however, this is the most commonly used structure. Figures 7(a) and 7(b) show two common implementations of uniform field across a thin oxide tun-

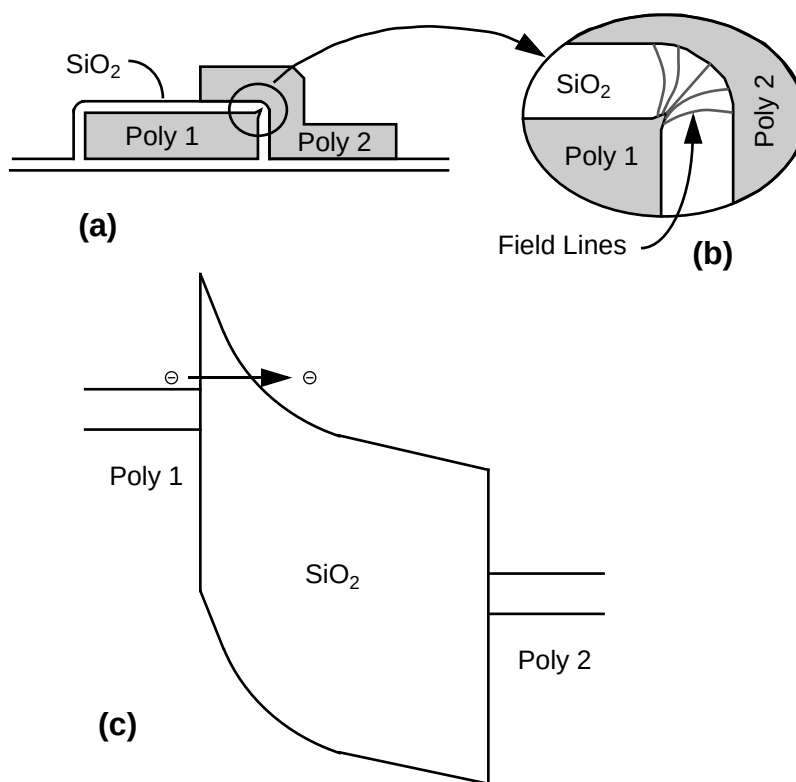


Figure 6—(a) Sketch of a cross-section through an enhanced tunneling emission cell; (b) enlargement of a portion of (a) that shows the concentration of field lines near the poly 1 corner; and (c) the resulting potential diagram in the vicinity of the corner illustrating how the high field concentration provides a locally thin barrier for enhanced tunneling.

nel area. Figure 7(a) shows a tunnel window that is opened through thicker oxide over a heavily doped n area, such as is typically employed in byte-alterable E²PROMs. Figure 7(b) illustrates the case in which the entire channel region is thin oxide, and a uniform field is applied between the gate and well. It is also possible to erase a floating gate by applying a high positive voltage between the one of the source/drain junctions and the gate, as is shown in Figure 7(c). This latter case necessarily results in fairly high lateral fields in the silicon, with consequences that will be discussed in more detail in the following subclause.

1.3.3 Band-to-band tunneling

In thin gate oxide MOS devices, a leakage current component can flow across the gated-diode surface p-n junction at a drain (or source) voltage well below the junction breakdown voltage, as shown in Figure 8 (see Parke, et al. [B86]). This leakage current typically shows a roughly exponential dependence on the voltage difference between the gate and the n⁺ diffusion; hence, it is often called the gate-induced drain leakage current (GIDL). As can be seen from Figure 8, the current can be changed by an order of magnitude at any given gate-drain bias by the junction profile.

This current has been identified as originating from the band-to-band tunneling process that occurs in the deeply depleted n⁺ surface region under the gate-to-drain (source) overlap area, as illustrated in Figure 9. Under the conditions in which the vertical field at the silicon surface is high (>0.8 MV/cm) and the surface potential drop is larger than the silicon band gap energy (1.12 eV), electron-hole pairs are generated by the tunneling of electrons from the silicon valence band into the conduction band. This process occurs predomi-

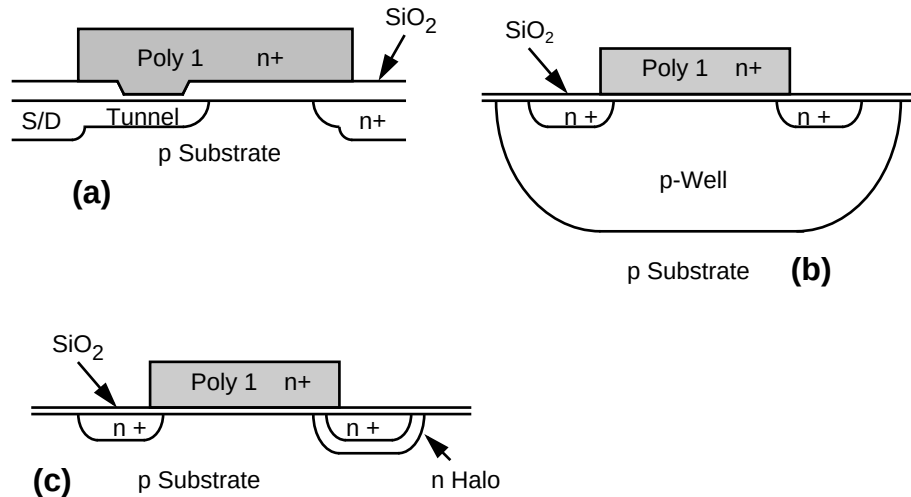


Figure 7—(a) A cross-section through a typical byte-alterable E²PROM with the thin oxide tunnel window over the n doped tunnel diffusion; (b) a cell with a p-well to allow the gate to be erased by biasing the well positive with respect to the gate poly, thereby causing tunneling across the thin oxide along the entire channel; and (c) a cell with an n-type halo around one diffusion to allow the floating gate to be erased by applying positive bias to the diffusion with respect to the gate

nantly in an area close to the p–n junction, where the surface doping of the n+ region is $\sim 5 \times 10^{18}$ atoms/cm³. Under these bias conditions, the generated holes flow to the p-type substrate, thereby maintaining the deep depletion condition, while the corresponding electrons flow to the n+ junction. Because this current is produced by the band-to-band tunneling process, it is referred to as the band-to-band tunneling current. Therefore, this current exhibits only a weak dependence on temperature.

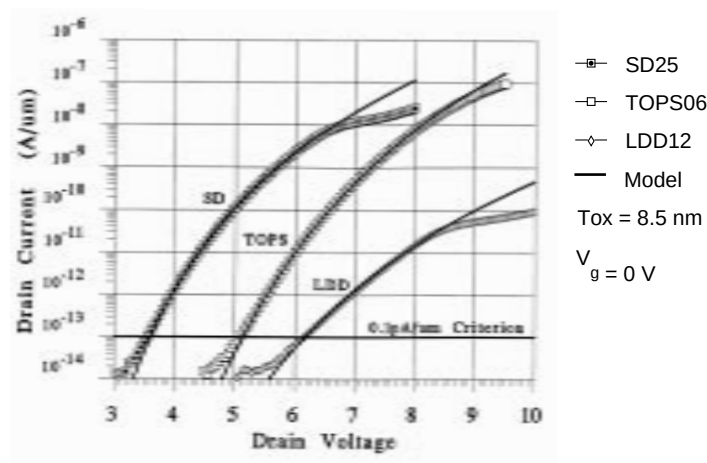


Figure 8—Drain current as a function of drain voltage with the gate grounded for three different junction profiles. The SD profile is an abrupt arsenic junction; the LDD profile has a phosphorus lightly doped drain; and the TOPS profile is an intermediate case (see Parke, et al. [B86]).

During the “edge” erase operation of flash EEPROM memory cells, such as the structure shown in Figure 7(c), the high field across the thin oxide induces an electron current, I_g , from the floating gate to the n+ diffusion caused by the F-N tunneling process. Because of the presence of a high vertical field in the silicon as well, a band-to-band tunneling process is initiated, as is shown in Figure 9, and a corresponding current flows across the p–n junction close to the silicon surface (see Chan, et al. [B12]). Generally speaking, this leakage current does not serve any useful purpose and should be minimized from both a performance and a reliability standpoint. In particular, the holes that flow through the surface depletion region are heated by the electric field across the junction and become “hot.” Some of these hot holes are injected into the tunnel oxide (see Chang and Lien [B13], Endoh, et al. [B24], Igura, et al. [B41], San and Ma [B95], Yoshikawa, et al. [B122]) and become trapped there. Trapped holes in the oxide have been shown to affect the data retention of the memory cell (see Haddad, et al. [B34], Verma and Mielke [B112]), to increase gate disturb effects in the memory array (see Brand, et al. [B7], Haddad, et al. [B33] and [B34]), to create electron traps in the oxide (see Tsuchiya [B105]), and to cause erratic erase behavior of the memory cell (see Ong, et al. [B82]). Holes trapped in the oxide also increase the tunneling current and speed up the erase operation, which leads to non-uniformities in the erased threshold voltage (see Ong, et al. [B82]).

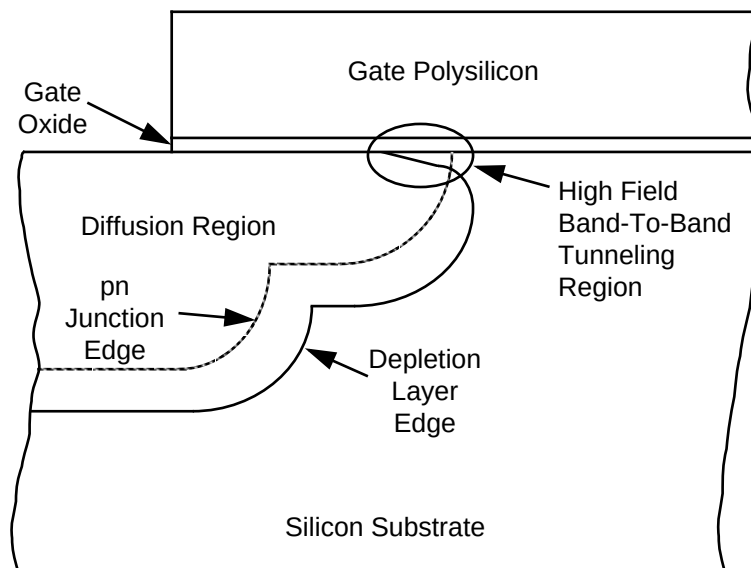


Figure 9—Illustration showing that when there is a high gate to drain bias, a depletion region can be induced in the diffusion region. The region in which the high local field may induce band to band tunneling is indicated.

Several solutions have already been proposed (see San, et al. [B95]).

- The use of a graded (e.g., double diffused) junction reduces the band-to-band tunneling current as well as the hot hole generation. There is, however, a trade-off in device design, since the graded junction has a large under-diffusion that increases the source (or drain) coupling to the floating gate and that prohibits aggressive channel length scaling.
- The so-called negative-gate-bias erase, where the voltage applied directly to the junction is greatly reduced (e.g., to 5 V), results in a much smaller hot hole generation rate and, thereby, realizes

increased reliability. The band-to-band tunneling current itself, however, is not decreased in this erase scheme.

- c) The channel erase scheme creates a high tunneling field from a negative gate bias only. This completely eliminates the band-to-band tunneling process, but has some considerable disadvantages. A higher (more negative) gate voltage is required for the same erase speed. Charge is trapped uniformly along the channel during erase, which can result in reduced read current. Triple well technology is used frequently to effectively implement this scheme (see Jinbo, et al. [B48]).

In the case of a graded junction, the band-to-band tunneling current is practically independent of the gate-to-junction voltage, and only drops to zero for junction voltages below 2 V. This suggests that a low junction voltage offers the best reliability conditions with respect to hole injection (see San, et al. [B95]).

1.3.4 Ultraviolet light erase

Light (photons) can be absorbed by matter and, in the process, can excite electrons in the matter to a high enough energy level that the electrons can surmount the barrier that binds them to the matter and can be emitted. Einstein won the Nobel prize for explaining that the energy that electrons acquire depends on the energy of the individual photon ($E = h\nu = h/\lambda$) and not on the intensity of the light (i.e., the number of photons). Thus, if the wavelength is longer than the minimum that is necessary for emission, no emission will occur. If the light is of sufficient energy to cause emission, the number of electrons emitted will be proportional to the number of photons (the light intensity).

Although Einstein's work was related to emission from metals into a vacuum, the same principles apply to the emission of electrons from silicon over the barrier of surrounding SiO_2 . As is illustrated in Figure 10, electrons can be excited from either the valence band or from the conduction band. Excitation from the conduction band requires a lower energy, $h\nu_1$, than excitation from the valence band, $h\nu_2$. However, excitation can only occur from the conduction band if there are electrons present to excite (i.e., the material must be doped fairly heavily n-type). Moreover, longer wavelength light that will excite electrons from the conduction band to above the oxide barrier can penetrate further into the silicon before being absorbed than shorter wavelength light. Thus, there is a good chance that the excited electrons will lose energy while traveling from the site of excitation to the Si/ SiO_2 interface. The consequence of this is that the chance for an electron excited from the conduction band to surmount the barrier is much less than that for an electron from the valence band. These features can all be seen in the data that is shown in Figure 11.

Several factors complicate this picture. Usually, the UV erasure takes place with the device unpowered so that photocurrents can flow in both directions (i.e., to both the control gate and the substrate). The Fermi levels have been omitted from Figure 10. Modern EPROMs are overwhelmingly n-channel devices, with the two gates doped n-type. However, the substrate is n-type in the source-drain overlap regions and p-type along the channel. Consequently, the Fermi level diagram is location dependent.

Another potential complication, shown in Figure 10, is that there is SiO_2 between both the floating gate and the control gate and between the floating gate and the channel. In practice, the dielectric between the two gates is often an oxide/nitride/oxide sandwich that acts to trap the photocurrent that is emitted in the direction of the control gate. This trapped charge inhibits further emission in the control gate direction. It can also later cause threshold shift of a programmed cell as the trapped charge redistributes (see Mori, et al. [B78] and Wu, et al. [B118]).

A final complication is the obstacles that the light may face in reaching the interfaces at which charge transport can occur. Because the silicon is a good absorber of light in the wavelengths of interest, the gates tend to shadow the effective regions. There is evidence that the light depends on both reflection and diffraction effects to slip in under the gates, and that this occurs primarily in the field oxide regions (see Katznelson and Frohman-Bentchkowsky [B52]). Some modern high-density technologies with multilevel metal interconnec-

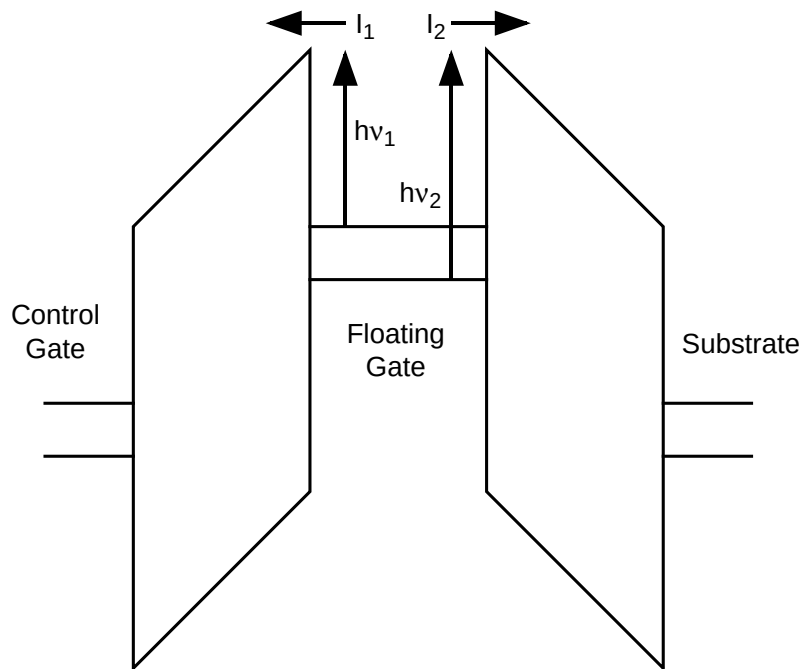


Figure 10—Schematic band diagram showing the excitation of electrons from the conduction band or from the valence band to energies at which they can surmount the oxide barrier and escape to the control gate or the substrate

tion layers may leave only a small fraction of the cell exposed to the light. This can cause rather slow erasure of these devices.

1.3.5 Read-out conditions

There are two different approaches to reading memory devices: random access and serial access. A random access memory is one in which any word in the address space can be accessed by selecting a unique address. Random access memories are exemplified by DRAMs and SRAMs. A semiconductor memory may have one or more parallel input/output (I/O) connections. Depending upon the number of I/O connections to the semiconductor memory and the width of the word in the system, the I/O connections of a number of semiconductor memories may be operated in parallel to supply the full word. A serial access memory is one in which a block of data is accessed in sequence. A serial access memory is exemplified by magnetic disks in which a track is selected and then the data on that track is accessed sequentially as the disk rotates under the read head.

Early UV-EPROMs were organized as random access memories with eight I/O connections so that a byte of information could be accessed by selecting an address. This is made easily possible by the array architecture of a typical UV-EPROM, which is illustrated in Figure 12(a). In this architecture, a single word-line selects a number of single transistor cells that have a common source connection. The state of any one of these cells can be queried by applying bias to its drain. This is referred to as a NOR architecture. Because random access is quite useful in memory that is used for code storage, the most common application for UV-EPROMs, this architecture is overwhelmingly dominant for this type of memory.

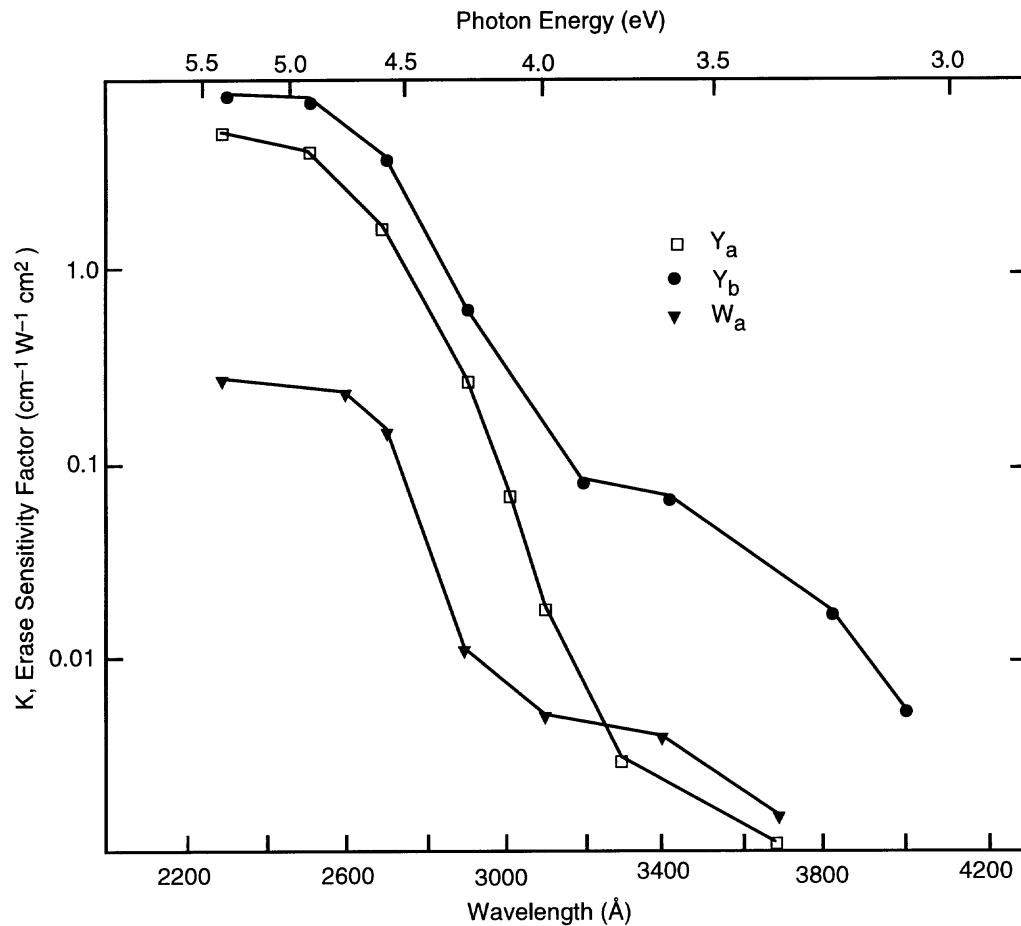


Figure 11—Spectral erase sensitivities of various UV-EPROM devices. Note that while all devices three show a large increase in sensitivity at a photon energy of ~ 4.2 eV, which corresponds to the energy to excite electrons from the top of the valence band to above the oxide barrier, the two cells with n-type floating gates, Y_b and W_a , also have weak shoulders extending to about 3.2 eV, which is the energy that is required to excite a conduction band electron over the barrier (see Katznelson and Frohman-Bentchkowsky [B52]).

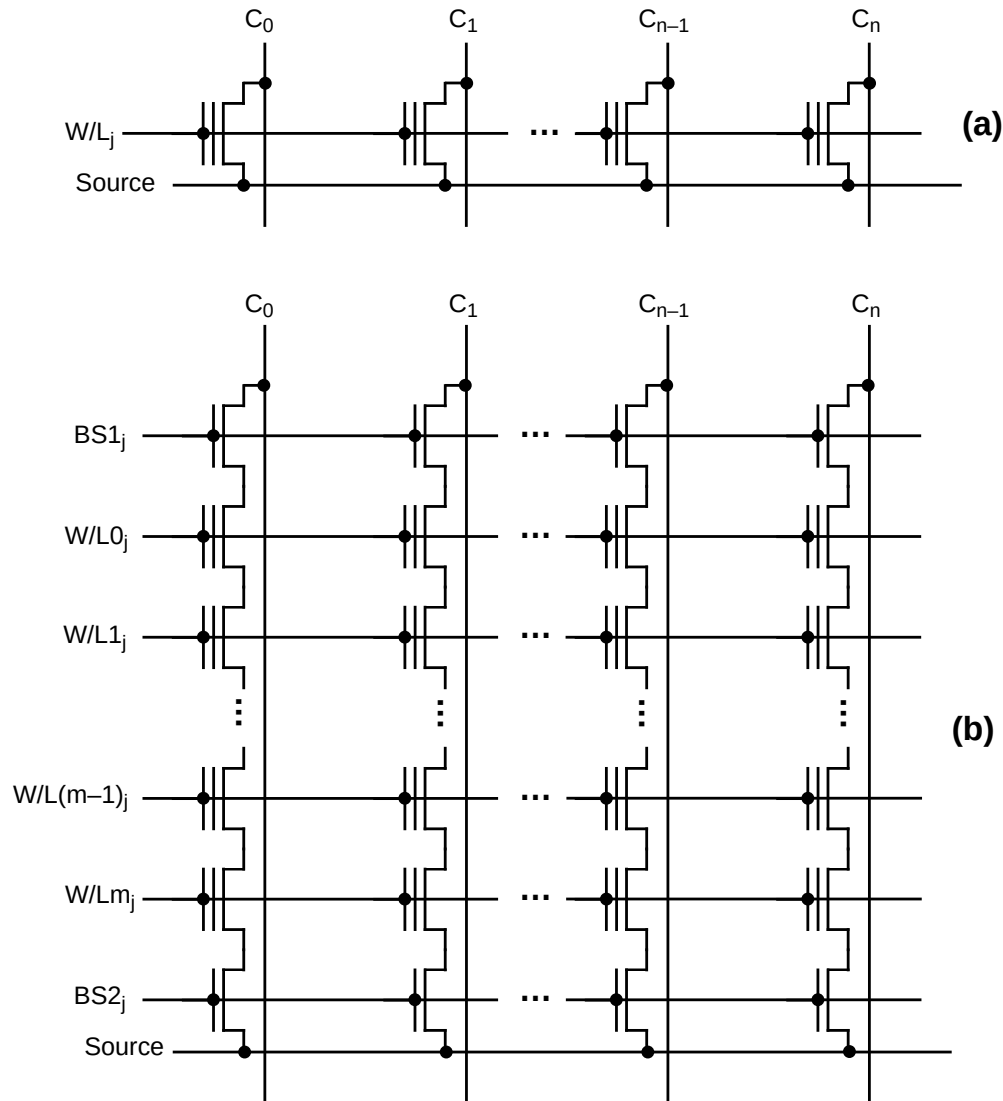


Figure 12—(a) Schematic of a row of floating gate cells arranged in a NOR architecture; and (b) schematic of a row of floating gate cells arranged in a NAND architecture

Byte-alterable E²PROMs also usually employ the NOR architecture and are frequently produced as random access memories. These memories, however, are also frequently used in applications in which a serial data bus is employed for cost reasons. In these applications, the information stored in a row is sensed in parallel and is transferred to a shift register that is then clocked to stream the data out in serial fashion. It is frequently possible in serial parts to randomly select the address of the first bit of the serial data stream, with the addresses being incremented automatically from the starting address.

Many flash EEPROMs employ the NOR architecture and provide random access to their data contents, especially those intended for program storage applications. Among those flash EEPROMs intended for mass information storage applications, sometimes referred to as disk replacement applications, however, serial data outputs are common. In some cases, the serial data is implemented with arrays that have a NOR archi-

tecture in the manner of byte-alterable E²PROMs. In other cases, a special architecture, usually referred to as the NAND architecture, is employed. This architecture is shown schematically in Figure 12(b). In the NAND architectures, as is seen in Figure 12(b), a number of floating gate transistors, usually 8 or 16, are arranged in series between the two select transistors in each cell. Each floating gate transistor acts as a storage element, but only one of the floating gate transistors can be accessed at any one time. Moreover, the read current through this cell is rather low ($\sim 1 \mu\text{A}$), so that the random access time is rather long (1–10 μs). Because of these factors, NAND memories are arranged as serial memories in which the data from a row is transferred in parallel into a set of latches from which they are read serially. Once the data are transferred to the latches, the data transfer rate can be $>10 \text{ MHz}$.

1.3.5.1 Binary sensing

The most common read condition for a floating gate memory is to couple voltage to the floating gate, such that the surface under the gate will be inverted for one charge state of the floating gate and will not be inverted for the other state. If the surface under the gate is inverted, current can flow and can be sensed as one logical state. If the surface is not inverted, no current flows, which is sensed as the opposite state.

To make this a little more concrete, consider, for example, the simple self-aligned stacked gate structure that is shown in Figure 13(a). The potential on the floating gate can be changed by the voltage applied to the gate above it, which will be called the control gate. The change in potential on the floating gate that results from a change in the voltage on the control gate is given by

$$\Delta V_{fg} = \Delta V_{cg} \times C_{cg_fg} / C_{tot} \quad (1)$$

where

- ΔV_{fg} is the change in floating gate potential;
- ΔV_{cg} is the change in control gate voltage;
- C_{cg_fg} is the capacitance between the floating gate and the control gate; and
- C_{tot} is the total capacitance between the floating gate and all nodes.

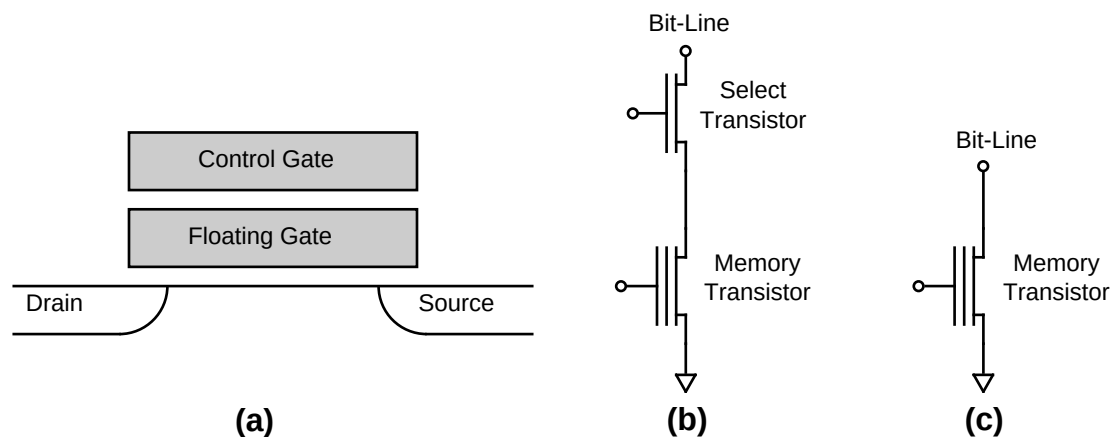


Figure 13—(a) Cross-section through a stacked self-aligned floating gate memory transistor, (b) schematic of the memory transistor in a cell with a select transistor, and (c) schematic of the memory transistor in a cell without a select transistor

The last factor is, in fact, the control gate coupling ratio. It is generally desirable to have this coupling ratio be greater than 0.5 in order for the voltage on the control gate to control the voltage on the floating gate. It is convenient to define the threshold of the floating gate transistor as that voltage that, when applied to the control gate, causes the floating gate transistor to conduct a certain amount of current with a given bias on the drain.

If the cell has a select transistor, as is illustrated schematically in Figure 13(b), it is enough that the floating gate threshold be above the high level of bias for the control gate for the programmed state and below this value for the erased state, as is shown in Figure 14(a). Because the series select transistor blocks current flow through an unselected cell, a dc bias can be applied to the control gate of the memory transistor so that erased cells all have thresholds of less than V_{SENS} and the programmed cells all have thresholds of greater than V_{SENS} . A convenient value for the control gate read bias is V_{SS} , but other values may be selected to properly center the threshold distribution. Sensing is accomplished typically by applying a small bias on the bit-line and by applying V_{DD} to the gate of the select transistor. If current is sensed, the cell is erased. If no current is sensed, the cell is programmed.

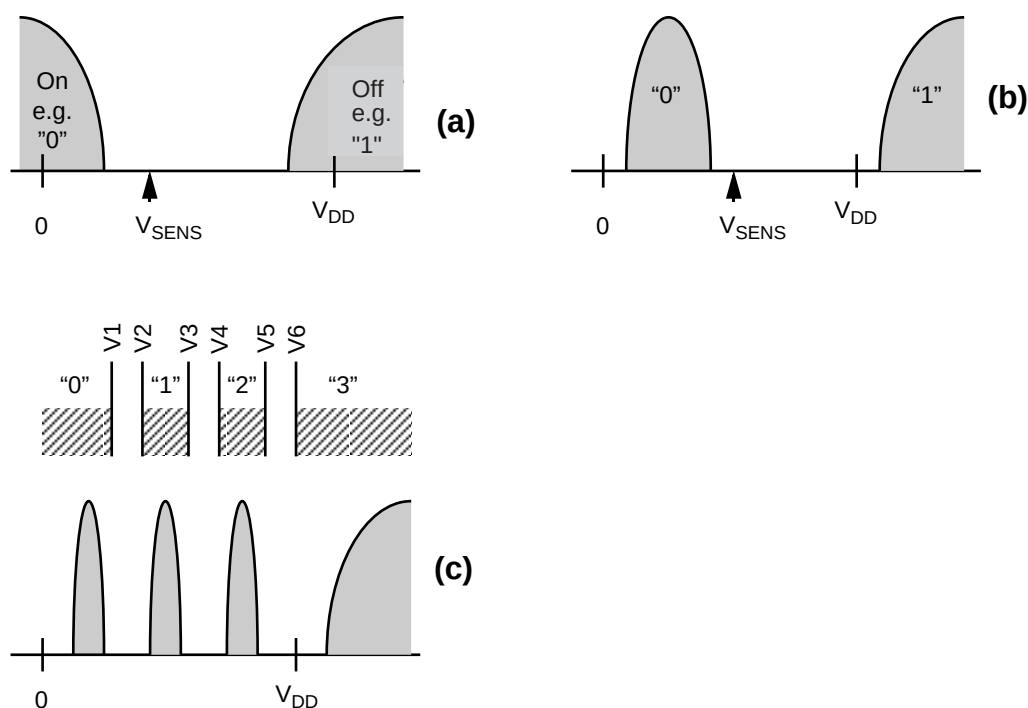


Figure 14—Illustrating typical threshold voltage ranges for (a) a single bit memory cell with select transistor, (b) a single bit memory cell without select transistor, and (c) two bit memory without select transistor

If the cell lacks a select transistor, as is shown in Figure 13(c), the bias on the control gate must serve to select the cell. The programmed cell threshold must be above the high-level level of the control gate. The erased cell threshold must be somewhat above 0 V, but must still be below V_{SENS} , as is shown in Figure 14(b). When these conditions are met, the floating gate will not conduct with the control gate at ground for either the programmed or erased state. This acts to deselect the cell. When the control gate is biased at its sensing voltage, the erased cell will conduct, but the programmed cell will not, thus allowing the charge state of the floating gate to be sensed.

1.3.5.2 Multilevel storage

If the threshold of the floating gate could be controlled so that it fell into a series of four bands with gaps in between, as is illustrated in Figure 14(c), it would be possible to determine into which of the four bands the threshold fell by applying multiple biases on the control gate. This would allow two bits to be stored in a single cell (Bauer, et al. [B4]). (Alternatively, the cell current could be sensed and divided into four groups.) As will be discussed in later clauses, there are a number of phenomena that will tend to affect the charge on the floating gate (and, hence, the cell threshold). The smaller the gap between the threshold values that constitute a state, the more difficult it becomes both to sense the state correctly and to maintain the threshold in the desired range. The potential gains in bit storage density, however, make this an attractive approach.

1.4 Introduction to floating gate memory arrays

The floating devices that are discussed in this standard all depend on charge storage on the floating gate for data storage and share many common concerns. Because of differences in application and in the development histories of the technologies, however, there are some significant differences in the manner in which these floating devices are integrated into arrays. These differences principally appear in the writing implementations.

UV-EPROMs principally are used for code storage and typically are only programmed once for the application. In fact, the significant majority of EPROMs sold today lack a window that would allow UV erase to occur and are sold as “one time programmable” memories. Programming is typically accomplished by means of a computer-controlled programming box. To match the application environment, the goal is to minimize the cost of the memory on a cost-per-bit basis. This results in the management of any required programming algorithms by the programming box. The programming box also supplies high-voltage pulses to power the programming function. As with many nonvolatile devices, EPROMs usually have an identification function that is integrated onto the chip so that the programming box can query a memory as to its identity in order to ascertain which programming algorithm to use. It is also common for EPROMs to generate separate control gate and drain programming voltage levels from the single high-voltage programming pulse.

Flash EEPROMs have a minimum erase block size that is greater than the minimum read block size. Byte-alterable E²PROMs have equal minimum erase and read block sizes.

Byte-alterable floating gate arrays generally program and erase via F-N tunneling, and may require that the write pulses be of the order of 20 V. Moreover, reliability is enhanced if the pulse is shaped to reduce the instantaneous tunnel current spikes. Early memory arrays left to the user the task of supplying these pulses, just as early RAMs lacked on-board sense amplifiers. All modern commercial arrays of this type, however, generate the necessary write pulses on-board the chip from a standard logic level power supply under the control of external logic level signals.

Flash EEPROM arrays fall into two types from the point of view of the application: those that allow writing from a single power supply voltage, which is also used as the system supply for read operation, and those that require two supply voltages during the write operation. The commercial market was initially dominated by the two supply versions. With time, however, the balance is shifting, and the single supply voltage approach is becoming dominant.

Because the flash EEPROMs are often intended for data alteration in the end application, it is important that the complexity demands on the system remain at a low level. To this end, the large majority of commercial flash EEPROMs integrate all the necessary control of the write procedure onto the IC, along with the array. Although these control circuits are often moderately complex state machines, the arrays that they accompany are of sufficiently high bit density that the fractional area of the chip that is devoted to the control circuitry remains in the range of a few percent.

In-system alterable arrays (i.e., EEPROMs), whether byte alterable or flash, are available with both random and serial data access. The data access mode is selected to match the application requirements.

Typical system bus clock frequencies are in the order of 5–50 MHz. Random access floating gate memory arrays are made with read access times that are compatible with system timing requirements. In order to avoid over stressing the chip components, however, it is desirable to allow writing to take place over a much longer time. To match the write times with the system cycle times, today's memory arrays control the long internal timings on-chip in response to logic signals with timings that are comparable to read access times.

Many other functions have been incorporated into some devices to make the in-system alterable arrays easier to use such as the ability to write many bytes per write cycle, the ability to detect the end of the internal write cycle, and the ability to inhibit writing when desired. These are discussed in more detail in Clause 9, along with the methods for testing them.

2. References

The following standards shall be used, when applicable, in preparing manuscripts. When the following standards are superseded by an approved revision, the revision shall apply.

ASTM F1192-88, Standard Guide for the Measurement of Single Event Phenomena from Heavy Ion Irradiation of Semiconductor Devices, Dec. 1988.¹

EIA/JESD21-C, Configurations for Solid-State Memories.²

EIA/JESD100-A, Terms, Definitions, and Letter Symbols for Microcomputers, Microprocessors, and Memory Integrated Circuits.

IEEE Std 100-1996, IEEE Standard Dictionary of Electrical and Electronics Terms, Sixth Edition.³

JEDEC 13.4, Test Procedure for the Measurement of Single Event Phenomena in Semiconductor Devices from Heavy Ion Irradiation, Preliminary Draft, Jan. 1992.⁴

MIL-STD 105D, Sampling Procedures and Tables for Sampling by Attributes.⁵

MIL-STD 414, Sampling Procedure and Tables for Inspection by Percent Defective.

MIL-STD 883C, Test Methods and Procedures for Microelectronics.

¹ASTM publications are available from the American Society for Testing and Materials, 100 Barr Harbor Drive, West Conshohocken, PA 19428-2959, USA.

²EIA publications are available from Global Engineering, 1990 M Street NW, Suite 400, Washington, DC, 20036, USA.

³IEEE publications are available from the Institute of Electrical and Electronics Engineers, 445 Hoes Lane, P.O. Box 1331, Piscataway, NJ 08855-1331, USA.

⁴JEDEC publications are available from JEDEC, 2001 I Street NW, Washington, DC 20006, USA.

⁵MIL publications are available from Customer Service, Defense Printing Service, 700 Robbins Ave., Bldg. 4D, Philadelphia, PA 19111-5094.

3. Definitions, abbreviations and acronyms, and symbols

3.1 Definitions

3.1.1 address: A character or group of characters that identifies a register, a particular part of storage, or some other data source or destination (see EIA/JESD100-A).

3.1.2 array: A group of memory cells that are arranged in a pattern.

3.1.3 array source: A common diffusion that provides the source of electrons for the cell in the read mode or, in the case of a cell that programs via channel hot electrons (CHE), provides the source of electrons for programming.

3.1.4 average outgoing quality limit (AOQL): A statistical measure of outgoing quality.

3.1.5 bit: **(a)** An abbreviation of binary digit; **(b)** a single occurrence of a character in a language that employs exactly two distinct kinds of characters; **(c)** a unit of storage capacity. The capacity, in bits, of a storage device is the logarithm to the base two of the number of possible states of the device (see IEEE Std 100-1996).

3.1.6 bit-line: The line that connects the memory cell drain to the sense amplifier during the read cycle and to a data line or latch during a write cycle.

3.1.7 bit-retention: The time interval from the writing of either state into a specified memory location to the earliest appearance of an incorrect state from such memory location, as measured on the specified output terminal(s).

3.1.8 block: A group of data that is contiguous in nature (see IEEE Std 100-1996). As used in this standard, it is understood to imply a subset of an entire memory array. *Syn:* **sector**.

3.1.9 block clear: Operation that sets a block of data within a memory to a common “1” state without affecting any other memory block.

NOTE—In the field of nonvolatile memories, clear conventionally means “set to a ‘1’ state.” This usage is adhered to in this standard.

3.1.10 block erase: **(a)** Signal (command) that causes the erasing of a block of data within a memory without affecting any other memory block; **(b)** the operation of erasing a block of data.

3.1.11 block select: An input terminal to which a signal must be applied in order to permit the device to read or write a block of data.

3.1.12 bulk erase: Operation of removing electrons from all of the bits of an array.

3.1.13 byte: A binary bit string that is operated on as a unit and is usually eight bits long and capable of holding one character in the local character set (see IEEE Std 100-1996).

3.1.14 byte clear: Operation that sets all bits in an addressed byte of memory to a common “1” state.

3.1.15 byte select transistor: The transistor, controlled by the word-line or row select line, that isolates each byte from the other bytes along the control line (each row of bytes).

3.1.16 cell: *See:* **memory cell**.

3.1.17 channel hot electrons (CHE): Electrons that are generated by sufficiently high electric fields in the channel with energies that exceed the thermal equilibrium energy.

3.1.18 charge pump: Circuitry that is used to create an on-chip voltage that is greater in magnitude than the voltage available from the device power supply. This voltage is typically used for the write operation.

3.1.19 chip: A small unpackaged functional element made by subdividing a wafer of semiconductor material. Sometimes referred to as a die (see IEEE Std 100-1996). Also used as a modifier to indicate an operation that applies to the entire chip as in chip enable or chip clear. *Syn:* **integrated circuit (IC).**

3.1.20 chip clear: **(A)** Operation that causes the writing of all memory elements to a common “1” state; **(b)** terminal to which the clear signal is applied (preferred usage).

3.1.21 chip erase: *See:* **bulk erase.**

3.1.22 chunk: A block of memory, typically 64 bits.

3.1.23 clear: To preset a storage or memory device to a prescribed state, usually that denoting zero (see IEEE Std 100-1996).

NOTE—In the field of nonvolatile memories, “clear” conventionally means to set the outputs of the memory to the high logic level (see EIA/JESD100-A).

3.1.24 clear algorithm: The timed sequence of signals necessary to clear the memory for a flash electrically erasable programmable read-only memory (EEPROM) (see EIA/JESD100-A).

3.1.25 clear disturb: The corruption of data in one location caused by the clearing of data at another location (see EIA/JESD100-A).

3.1.26 column select line: The line that is determined by the column addresses (output of the Y decoder) that are used to select the appropriate access transistors during a read or write.

3.1.27 column select transistor: The transistor, controlled by the column select line, that accesses the appropriate bit-line during a read or write cycle.

3.1.28 control capacitor: The element by which voltage is coupled to the floating gate for reading or writing.

3.1.29 control gate: A form of the control capacitor in which the top polysilicon gate, above the floating gate, of the memory transistor is connected to the word-line or control line.

3.1.30 control line: The line, connected to the memory transistor control element, that provides the reference voltage to the memory cell during a read and may provide a high voltage during a write cycle.

3.1.31 data bar polling: An end-of-write indicator. (See EIA/JESD21-C for a complete description.)

3.1.32 data change: An event in which at least one bit of data is caused to change (see EIA/JESD100-A).

NOTE—This event may be used as a unit of endurance for erasable programmable read-only memories.

3.1.33 data cycle: A cycle in which each bit changes to its opposite state and back to its original state (see EIA/JESD100-A).

NOTE 1—These changes may occur for all bits in parallel or series, e.g., by page, block, word, byte, or bit.

NOTE 2—This cycle may be used as a unit of endurance for erasable programmable read-only memories.

3.1.34 data latch: Latch to temporarily store data to be subsequently written into a floating gate memory.

3.1.35 data line: *See:* **bit-line.**

3.1.36 data retention time: *See: retention time.*

3.1.37 data rewrite: An operation in which data is written into any array and that includes one data cycle or at least one data change (see EIA/JESD100-A).

3.1.38 dc erase disturb: Sometimes called deprogramming or gate disturb. Floating gate to control gate charge loss when the word-line (and control line) is high for programming another location on the same word-line (and control line).

3.1.39 dc program disturb: Sometimes called word-line disturb. Floating gate charge gain when the word-line (and control line) is high for programming another location on the same word-line (and control line).

3.1.40 dc static tests: Voltage and current tests to determine the input levels, output levels, and power dissipation characteristics of a device.

3.1.41 dose rate: The time rate of deposition of ionizing radiation.

3.1.42 drain erase disturb: Floating gate to drain charge loss, on an erased cell, when the bit-line is high for programming or erasing another location on the same bit-line.

3.1.43 dynamic radiation test: Test of the instantaneous effects of radiation that are obtained by monitoring electrical and memory properties of interest continuously during and immediately after exposure to a transient radiation pulse.

3.1.44 EEPROM redundancy: *See: redundancy.*

3.1.45 electrically erasable programmable read-only memory (EEPROM): A reprogrammable read-only memory in which the cells at each address can be erased electrically and reprogrammed electrically (see EIA/JESD100-A).

NOTE—These devices are characterized by slower write times than read times. The memory is used for applications that specify a maximum number of write operations.

3.1.46 endurance: The ability of a reprogrammable read-only memory to withstand data rewrites and still comply with its specifications (see EIA/JESD100-A).

3.1.47 E²PROM: *See: electrically erasable programmable read-only memory (EEPROM).*

NOTE—In this standard, the convention has been adopted that E²PROM implies a byte-alterable memory, while EEPROM implies an array with only block or chip clear.

3.1.48 erasable programmable read-only memory (EPROM): A reprogrammable read-only memory in which all cells can be simultaneously erased using ultraviolet light and in which the cells at each address can be reprogrammed electrically (see EIA/JESD100-A).

3.1.49 erase: In the field of electrically erasable programmable read-only memories, the removal of electrons from the floating gate of the memory cell (see EIA/JESD100-A).

3.1.50 erase algorithm: The timed sequence of signals necessary to erase the memory for a flash electrically erasable programmable read-only memory (EEPROM) (see EIA/JESD100-A).

3.1.51 erase disturb: The corruption of data in one location caused by the erasing of data at another location (see EIA/JESD100-A).

3.1.52 erase margin: The minimum measured difference between the erased states and the sensing level for the array.

3.1.53 erase-program cycle: The event of writing a memory cell from the erased state to the programmed state and back to the erased state. *Contrast:* **program-erase cycle.**

NOTE—This event may be used as a unit of measurement for endurance. Within a sequence, erase-program cycles are indistinguishable from program-erase cycles (see EIA/JESD100-A).

3.1.54 failures in time (FIT): A statistical measure of failure rate corresponding to one failure in 10^9 hours of device operation.

3.1.55 false-write protection: *See:* **inadvertent-write protection.**

3.1.56 FAMOS[®]: Floating gate avalanche metal-oxide semiconductor (MOS) transistor used in erasable programmable read-only memory (EPROM).⁶

3.1.57 flash EEPROM: An electrically erasable programmable read-only memory (EEPROM) in which clearing can be performed only on blocks or the entire array (see EIA/JESD100-A).

3.1.58 Fowler-Nordheim (F-N) tunneling: A quantum-mechanical effect in which electrons penetrate through a barrier region in which they have no allowed states and emerge in the conduction band of the barrier as a result of an externally applied electric field.

3.1.59 functional dynamic tests: Operation of the array at specified dynamic conditions to determine functional performance.

3.1.60 hot electron programming: The injection of energetic “hot” electrons onto the floating gate.

3.1.61 inadvertent write protection: The circuitry that is used to prevent writing from occurring when the control signals of the memory enter an uncontrolled state during power on, power off, or noise transients (see EIA/JESD100-A).

3.1.62 incremental total test dosage: Test of permanent changes induced by radiation that are obtained by a comparison of characteristics before and after exposure to a given integrated total dose level.

3.1.63 insulated-gate field-effect transistor (IGFET): A field-effect transistor that has one or more gate electrodes that are electrically insulated from the channel.

3.1.64 integrated circuit (IC): A combination of interconnected circuit elements inseparably associated on or within a continuous substrate (see IEEE Std 100-1996). *See also:* **chip.**

3.1.65 latch-up: A state in which a low-impedance path results from an input, output, or supply excessive operating condition that triggers a parasitic structure and persists after the removal or cessation of the triggering condition.

3.1.66 linear energy transfer (LET): The amount of energy that is transferred from an ionized particle to the medium through which it passes per normalized unit of distance traveled.

NOTE—Usually expressed in units of $\text{MeV} \times \text{cm}^2/\text{mg}$.

3.1.67 mass erase: *See:* **bulk erase.**

3.1.68 memory array: *See:* **array.**

3.1.69 memory capacity: *See:* **capacity.**

⁶FAMOS is a registered trademark of Intel Corp.

3.1.70 memory cell: The combination of one or more single or merged transistors formed to provide a means of accessing, changing, and storing data.

3.1.71 memory integrated circuit: An integrated circuit consisting of memory cells and usually including associated circuits such as signal amplification and address selection (see EIA/JESD100-A).

3.1.72 memory location: A subdivision of a memory, including one or several memory cells, that is the smallest part of the memory that can be addressed.

NOTE—The content of a memory location is usually called a bit, a byte, or a word, as appropriate (see EIA/JESD100-A).

3.1.73 memory organization: The arrangement of memory cells, either by geometrical arrangement in rows and columns or by organization of the data to be stored.

3.1.74 nonvolatile memory: A memory in which the data content is retained when power is no longer supplied to it (see EIA/JESD100-A).

3.1.75 one-transistor cell: A memory cell that is accessed within the physical confines of a single source and drain.

3.1.76 over-erase leakage: The current on the bit-line caused by over-erased cell(s).

3.1.77 over-erase recover: A custom programming algorithm to raise the threshold of an over-erased cell.

3.1.78 over-erased cell: An unselected cell that has excessive source-drain leakage current resulting from an erase operation that has reduced the threshold of the cell below the applied control gate voltage.

3.1.79 over-erased device: A device that cannot be read or programmed correctly because of excessive over-erase leakage.

3.1.80 page: A section of the array that may be written to simultaneously. This usually refers to more than one byte.

3.1.81 page mode: A method of operation in which more than one memory location is written simultaneously.

3.1.82 pass gate: The transistor(s), controlled by the word-line, to enable the programming of bits along the selected word-line and the isolating of bits on unselected word-lines from the array source.

3.1.83 polycrystalline silicon: A silicon layer that contains a multiplicity of crystals. This is the common form of silicon that is employed for transistor gates and interconnections on the surface of dielectric layers in MOS circuits.

3.1.84 power up/down protection: See: **inadvertent write protection.**

3.1.85 program: The operation of injecting electrons onto the floating gate of the memory cell (see EIA/JESD100-A).

3.1.86 program disturb: The corruption of data in one location caused by the programming of data at another location.

3.1.87 program-erase cycle: The event of writing a memory cell from the programmed state to the erased state and back to the programmed state (EIA/JESD100-A). See also: **erase-program cycle.**

NOTE—This event may be used as a unit of measurement for endurance. Within a sequence, program-erase cycles are indistinguishable from erase-program cycles.

3.1.88 program margin: The minimum measured difference between the programmed states and the sensing level for the array.

3.1.89 programming (write) algorithm: Typically for flash memories or ultraviolet-erasable programmable read-only memories (UV-EPPROMs), the specified timed sequence of signals and levels necessary to cause the device to program.

3.1.90 radiation-induced data loss characteristics (data upset): Stored data integrity as a function of total dose or dose rate after a given test pattern has been written into the device.

3.1.91 rads in Si: Amount of radiation measured by its ionizing effect in silicon; one rad equals 100 erg of energy deposited in a gram of irradiated solid (see IEEE Std 100-1996).

3.1.92 random-access memory (RAM): A memory that permits access to any of its address locations in any desired sequence with similar access time to each location (adapted from IEC 748-2).

NOTE—The term RAM, as commonly used, denotes a read/write memory with unlimited data rewrite capability and equal read and write times.

3.1.93 read circuitry: That part of the memory that is used in transferring the stored data from the memory section to external circuitry.

3.1.94 read disturb: The corruption of data that is caused by reading the memory (see EIA/JESD100-A).

3.1.95 read-only memory (ROM): A memory in which the contents are intended to be read only and not altered during normal operation.

3.1.96 ready/busy: An end-of-write indicator. (See EIA/JESD21-C for a complete description.)

3.1.97 redundancy: The provision of extra memory cells, usually rows or columns, that can be mapped into the memory array to replace defective cells.

NOTE—In nonvolatile memory, the mapping may be controlled through EEPROM or other fuse techniques (see EIA/JESD100-A).

3.1.98 retention time: The time interval between the instant that data is stored and the instant that the data can no longer be read correctly (see EIA/JESD100-A).

NOTE—Unless otherwise qualified, the term “read-only memory” implies that the data content is determined by the structure of the memory and is unalterable (see EIA/JESD100-A).

3.1.99 row select gate: Sometimes called the select gate. The transistor, controlled by the word-line, that isolates the memory transistor from the bit-line so that individual bytes may be altered.

3.1.100 row select line: The line, determined by the row addresses (output of the X decoder), that is used to access the appropriate rows and, when present, the byte select transistors during a read or write.

3.1.101 row select transistor: Sometimes called the select transistor or select gate. The transistor, controlled by the row select line, that isolate the memory transistor from the bit-line so that individual bits are isolated from voltages on the bit-line.

3.1.102 sector: See: **block**.

3.1.103 sector select line: The line, determined by the row addresses (output of the X decoder), that is used to access the appropriate sector select transistor.

3.1.104 sector select transistor: The transistor, controlled by the sector select line, that isolates the sector source from other sectors.

3.1.105 sense amplifier voltage window: Sense amplifier sensitivity between differential state or reference voltage and sensed state.

3.1.106 single-event upset (SEU): The loss of data that is caused by passage of a single ionizing particle through an array.

3.1.107 software data protection (SDP): A means of preventing inadvertent write or access to different operating modes. (See EIA/JESD21-C for a complete description.)

3.1.108 static dose rate test: Test of the permanent changes induced by radiation that are obtained by a comparison of characteristics before and after exposure at a given dose rate.

3.1.109 storage capacity: The amount of data that can be contained in a storage device (see IEEE Std 100-1996).

3.1.110 storage cell: An elementary unit of storage (e.g., a binary cell or a decimal cell) (see IEEE Std 100-1996).

3.1.111 timer EEPROM: Typically a byte-alterable electrically erasable programmable read-only memory (EEPROM) with on-chip latches for all address, data, and control lines to internally control the duration of a write cycle.

3.1.112 timing sequence: Sequence of enable, coding, and data pulses to permit writing or reading of information.

3.1.113 toggle bit: An end-of-write indicator. (See EIA/JESD21-C for a complete description.)

3.1.114 total dose: The total amount of ionizing radiation that is deposited in the active area of a device over a given period of time. The unit of measure for total dose that is most commonly used in the present context is rads(Si) (see IEEE Std 100-1996).

3.1.115 ultraviolet-erasable programmable read-only memory (UV-EPROM): *See: erasable programmable read-only memory (EPROM).*

3.1.116 window: The measured difference in voltage or current between the erased and programmed states.

3.1.117 word: A character string or bit string that is considered as an entity (see IEEE Std 100-1996).

3.1.118 word clear: Operation that sets all bits of a word to a common logic “1” state.

3.1.119 word erase: The operation of removing the electrons from all bits in a word.

3.1.120 word-line: The line, determined by the row addresses (output of the X decoder), that is used to access the appropriate memory transistors, pass gate, or byte and row select transistors during a read or write.

3.1.121 write: The operation by which a data state is entered into one or more memory cells. This may be a program or an erase operation.

3.1.122 write circuitry: Section of the memory that is used in the alteration of the stored data during the write operation.

3.1.123 write disturb: The corruption of data in one location caused by the writing of data at another location (see EIA/JESD100-A).

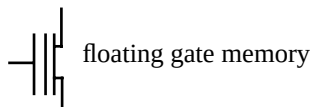
3.2 Acronyms and abbreviations

A	atoms
Å	angstrom (10^{-10} meter)
AOQL	average outgoing quality limit
ATE	automated test equipment
CAM	content addressable memory
C_G	gate capacitance per unit area
CHE	channel hot electrons
cm	centimeter (10^{-2} meter)
cm^2	square centimeter
cm^3	cubic centimeter
Δ	delta, difference between
ECC	error checking and correction
EELPD	electrically erasable programmable logic device
E^2PROM	electrically erasable programmable read-only memory
EEPROM	electrically erasable programmable read-only memory
EPROM	erasable programmable read-only memory
eV	electron volt (1.602×10^{-19} Joules)
FAMOS	floating gate avalanche MOS
FET	field-effect transistor
FIT	failures in time
FLOTOX [®]	floating gate tunnel oxide ⁷
F-N	Fowler-Nordheim
FXR	flash x-ray
λ	wavelength
galpat	galloping pattern
h	Plank's constant, 6.6262×10^{-34} Joules-sec
$\hbar$	$\hbar/2\pi$, 1.05459×10^{-34} Joules-sec
IC	integrated circuit
I_{DS}	drain-source current
IGFET	insulated-gate field-effect transistor
J	current density
k	Boltzmann's constant, 1.38062×10^{-23} Joules $\times$ °Kelvin ⁻¹
k	kilo (10^3)
K	1024, 2^{10}
k'	gain factor
l	length of channel in the direction of current flow
LET	linear energy transfer
LSI	large-scale integration
μ	carrier mobility
μm	micron, 10^{-6} meter
MNOS	metal-nitride-oxide semiconductor
MOS	metal-oxide semiconductor
μs	microsecond (10^{-6} second)
ms	millisecond (10^{-3} second)
MV	megavolt
v	frequency

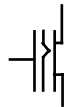
⁷FLOTOX is a registered trademark of Intel Corp.

n+	diffusion with excess donors
N_{end}	endurance
nm	nanometer (10^{-9} meters)
NOVRAM	nonvolatile random access memory
ns	nanosecond (10^{-9} second)
NVRAM	nonvolatile random-access memory
ONO	oxide/nitride/oxide
p+	diffusion with excess acceptors
poly	polycrystalline silicon
polysilicon	polycrystalline silicon
q	elementary charge, 1.60219×10^{-19} Coulombs
rad	100 erg/gram
RAM	random-access memory
ROM	read-only memory
SCSG	source coupled split-gate
SDP	software data protection
SEP	single-event phenomena
SEU	single-event upset
Si	silicon
SIGL	stress-induced gate leakage
SiO ₂	silicon dioxide
SNR	signal-to-noise ration
T	temperature
TPVG	triple poly, virtual ground
t_{ret}	retention time
UV-EPROM	ultraviolet-erasable programmable read-only memory
V	volt
V_{DS}	drain-source voltage
V_{GS}	gate-source voltage
V_{t}	threshold voltage
w	width of the channel
ϕ_{f}	Fermi potential

3.3 Symbols



floating gate memory



floating gate memory with a thin oxide region for enhanced tunneling conduction near the drain

4. Basic floating gate FET types

4.1 Background

The floating gate FET structure was first proposed by Kahng and Sze [B51] in 1967. The first commercial realization of a floating gate memory device was the floating-gate avalanche-injection MOS memory (FAMOS) device developed by Frohman-Bentchkowsky ([B29], [B30]) in the early 1970s. Programming of

the FAMOS device was accomplished by avalanche breakdown of the drain junction. Hot electrons generated by this process were transported to the floating gate. Hot electrons could overcome the oxide barrier with relatively low applied voltage. To erase the FAMOS memory, ultraviolet light was used to excite electrons over the oxide barrier. By the late 1970s, FAMOS devices had evolved into a generic memory product called electrically programmable read only memory (EPROM) that used CHE for programming (Salsbury et al. [B93]).

The obvious drawback of the UV-EPROM is that it required ultraviolet (UV) light for erasing. This requires a quartz window in the package and the removal of the memory component from the circuit board for insertion into a special UV eraser. Furthermore, a high voltage (12–21 V) and high current supply (>10 mA), in addition to V_{DD} , is required for programming. Consequently, instead of programming in-system, programmers are used for programming EPROMs.

Historically, the next development in floating gate memories occurred in 1981 with the introduction of the commercial electrically erasable programmable read only memory (EEPROM or E²PROM) that uses the tunneling of electrons onto and from the floating gate for programming and erasure (Johnson et al. [B49]). Because this is a relatively low current process, the necessary voltages for writing could be generated on-chip with charge pumps and a single supply. In-system byte-alterable E²PROM soon became a standard memory component that was available from a number of suppliers (Kianian, et al. [B55]; Gupta, et al. [B31]).

The principal drawback to the E²PROM technology for many applications is cost. The standard E²PROM cell contains two or more transistors that require more area than other cheaper memory technologies, such as ROM, EPROM, and DRAM that require only a single transistor. To address this challenge, a floating gate memory technology has been developed that abandons byte alterability in exchange for lower cost. This technology is called “flash memory.” As of the time of the writing this standard, there are a number of commercial available variants that are described herein.

All three of these floating gate memory technologies, EPROM, byte-alterable E²PROM, and flash EEPROM, are in current widespread use and will be described in this standard. The cell approaches that are currently in use are described in the following paragraphs, grouped by the general type (i.e., EPROM, E²PROM, or flash).

4.2 EPROM

4.2.1 General description

The basic floating gate EPROMs feature a polysilicon field plate, usually doped to degenerate n⁺ by phosphorus, that is insulated from the channel region of the transistor. This field plate is completely insulated from all other electrodes and silicon by high-quality insulators, thus forming a floating gate. Typically, there are one or more electrodes that are coupled capacitively to the floating gate. Electron conduction of the transistor channel can be controlled by the charge that is stored on the floating gate. All modern commercial EPROMs are n-channel devices, although the original FAMOS devices were p-channel (Salsbury et al. [B93]). Programming the cell injects excess electrons onto the floating gate, which increases the threshold for conduction. Erasure with ultraviolet light removes the excess electrons and restores the initial threshold, as is described in Clause 1.

4.2.2 T-cell UV-EPROM cell

4.2.2.1 Description

The T-cell is the cell that is used in the majority of all UV-EPROMs ever built. The name comes from the shape of the active region as seen in a cell layout. The cell is programmed by drain-side hot electron injection.

This cell usually employs self-aligned stacked polysilicon gates to achieve a compact cell that is suitable for a high bit density array.

The layout and cross-sections of the T-cell are shown in Figure 15. The T shape of the active region is clearly seen. The bit-line contact to the drain is shared between adjacent cells as is the source diffusion. The poly 2 word-line, connected to the control gate in Figure 15, extends through all of the cells in a row. In some cases, this word-line may be strapped in metal or a refractory metal silicide to reduce RC delay. It is also common to employ straps for the source line to reduce voltage drops in the source diffusion that are of special concern while programming. The floating gate is self-aligned (i.e., the edges of the floating gate are coincident with those of the overlying control gate in the direction of current flow).

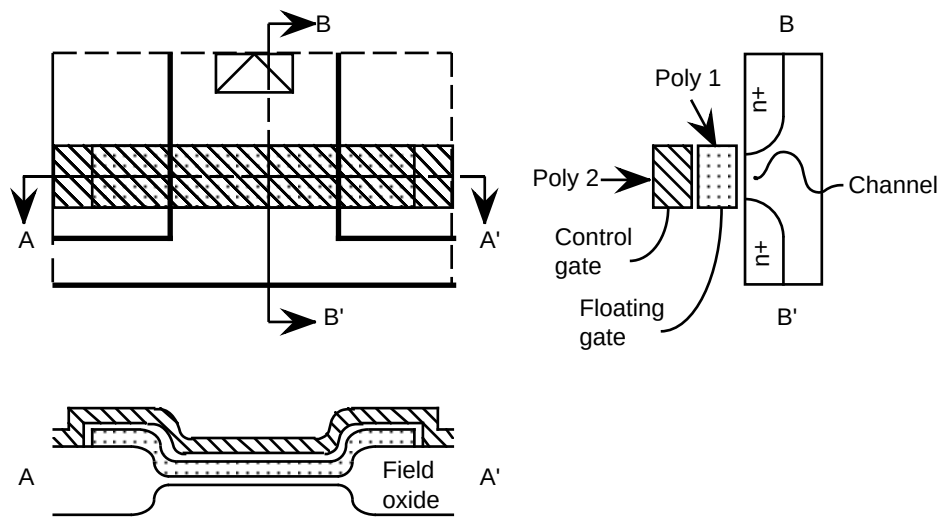


Figure 15—Top view and cross-section views through a typical T-cell EPROM with a self-aligned floating gate

The schematic drawing of a portion of an array in Figure 16 displays the interconnections of the cells. The control gates are connected by word-lines (W/L) in rows. The bit-lines (B/L) connect the drains in columns. Note that there are no separate select transistors in the array. This contributes to the small size of this cell.

4.2.2.2 Programming

Recall from the introduction that to program a cell with drain-side hot carrier injection, it is necessary to bias both the drain and the floating gate at moderately high voltages. The floating gate is completely encased in a high-quality dielectric in order to retain the charge that represents the information. Therefore, biases must be applied to the floating gate through capacitive coupling. There are capacitances from the floating gate to the polysilicon word-line, to the source, to the drain, to the channel, and to the substrate through the field oxide. Because it is desired to control the floating gate from the word-line, often called the control gate, the capacitance between these two lines should be relatively large. Arbitrarily enlarging the interpoly capacitance, however, would enlarge the cell; consequently, a compromise must be made. The usual result is that the interpoly capacitance is 50–60% of the total capacitance of the floating gate so that 50–60% of the word-line bias is coupled to the floating gate.

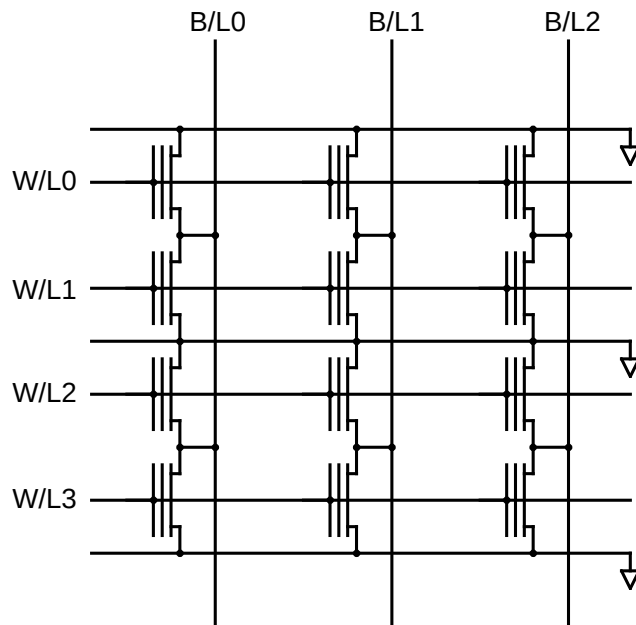


Figure 16—Schematic drawing of a 3×4 cell portion of an EPROM memory array

To program a typical cell, a programming voltage, V_{pp} , is applied to the selected word-line and $\sim 1/2 V_{pp}$ is applied to the drains of the transistors to be programmed. Because the peak drain programming current for drain-side programming is ~ 0.5 – 1 mA, the programming is usually done a byte at a time, although a few of the more recent high bit density EPROMs have featured multibyte programming as a way of reducing the total time to program an array.

4.2.2.3 Inhibit

The T-cell NOR arrays naturally inhibit spurious programming of the neighboring cells. The cells connected to the unselected bit-lines are usually left floating and are not programmed because no current flows. The cells on the unselected (grounded) word-lines should be nonconducting because the word-line coupling keeps the floating gates below threshold.

4.2.2.4 Read

Read is accomplished by applying a bias, typically V_{DD} , to the selected row and by applying a smaller positive bias to the bit-lines of the columns that are being interrogated. The floating gates of the erased cells are coupled above threshold by the word-line bias, and these cells consequently conduct. Programmed cells have enough negative charge on their floating gates that they remain below threshold in the presence of V_{DD} on their word-lines.

4.2.2.5 Disturbs

The T-cell is subject to program disturbs from the word-line on unselected rows and from the bit-line on unselected columns, as discussed in Clause 8. These disturbs tend to be defect induced, and often are screened during production testing. Two other possible disturbs deserve design attention. The read condition

differs from the program condition only in the magnitude of the applied voltages. The drain voltage during read must be low enough so that the total number of hot electrons that are collected on the floating gate of an erased cell in the read mode over the life of the part will be less than that causing the allowable threshold shift. This is typically of the order of one electron every few hours, a standard that requires a very low generation rate. It is also necessary that the floating gate be long enough that the programming voltage on the drain not punch-through the unselected transistors despite capacitive drain coupling that tends to turn on the floating gate transistor. Punch-through may result in the programming of an erased cell.

4.2.3 X-cell EPROM cell

4.2.3.1 Description

The X-cell is composed of four self-aligned gate EPROM elements in which the diffusion regions of a 2×2 bit matrix form an "X" shape. The X-cell is electrically connected as a virtual ground array in which the ground for an element is selected. All diffusions that are located in the same column are connected by metal. The advantage of the X-cell is a denser memory array than grounded source (T-cell) cells. The disadvantage is that the X-cell requires more complex decoding and design consideration of inhibit conditions that are created by parasitic conduction paths. Figure 17 shows a layout of a 2×2 bit matrix. Figures 18 and 19 show corresponding cross-sections.

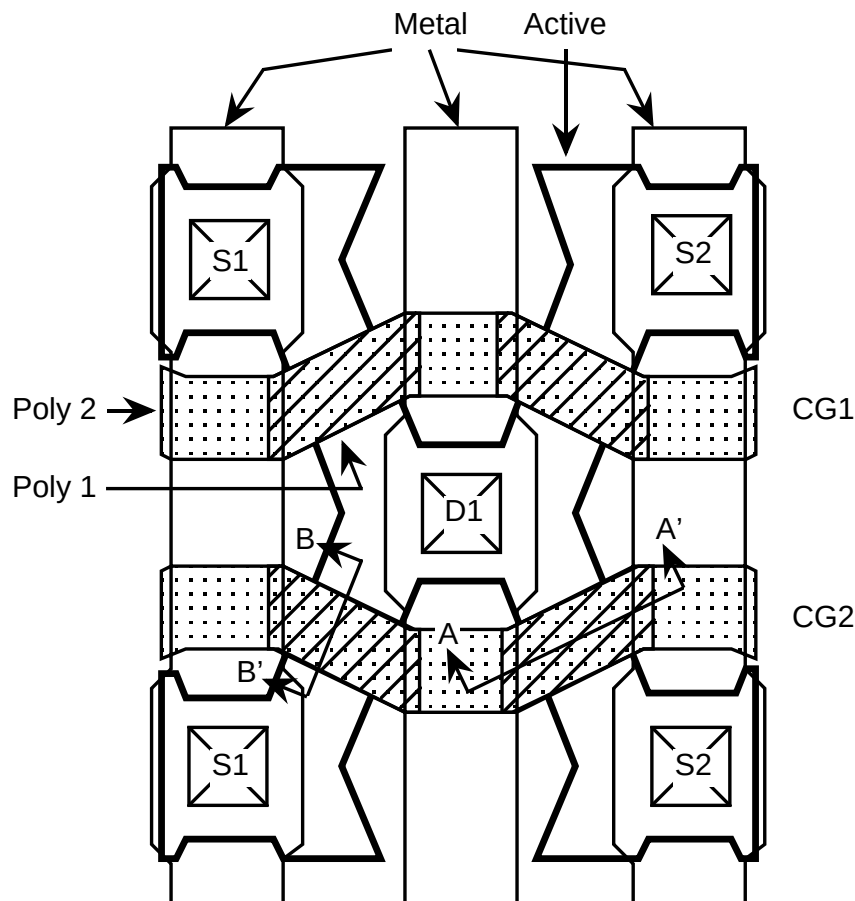


Figure 17—Top view of an X-cell layout (the active region is in the form of the letter X)

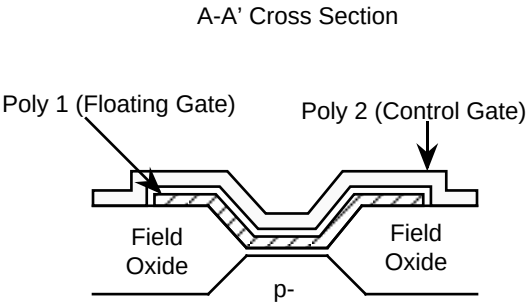


Figure 18—Section through the element along the line A–A'

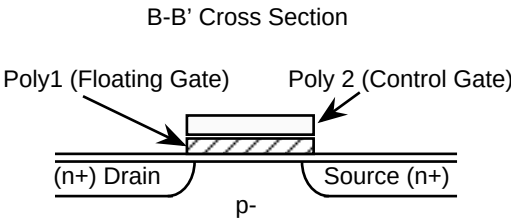


Figure 19—Section through the element along the line B–B'

4.2.3.2 Charge transfer

The potentials that are applied to an element to be programmed in an X-cell EPROM are the same as with the T-cell EPROMs. See Figure 20. Electrons are placed on the floating gate via CHE injection at the drain. This is achieved by placing high voltages on the drain and control gate and by holding the source near ground. Electrons are removed from the floating gate via UV erase.

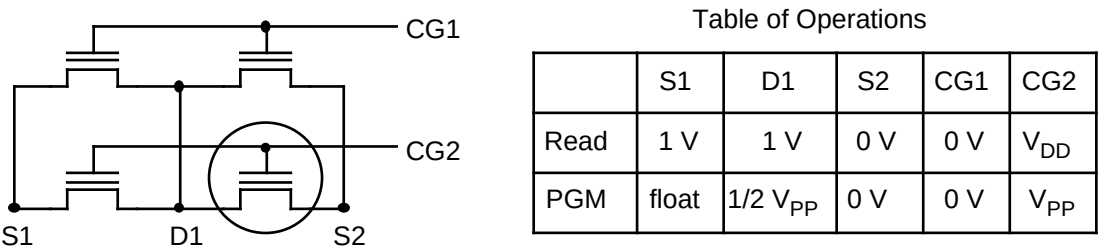


Figure 20—Schematic representation and table of operation for the X-cell matrix shown in Figure 17, cross-sections of the elements that are shown in Figures 18 and 19

4.2.3.3 Inhibit conditions

In a virtual ground configuration, elements that have common control gates form series connections of EPROM transistors. A charge may be transferred through these series connections, depending upon an element's V_t and source and control gate potentials. This creates a potential spurious programming condition

for an element that shares a common drain and is physically adjacent to the element that is being programmed. The adjacent element has its drain and control gate at the same potentials that are required for programming. Depending upon the adjacent element's source potential, unintentional drain to source currents may result in charge injection onto the adjacent element's floating gate. Proven design techniques avoid spurious programming of any element with high voltage on its drain by driving the element's source to its drain potential or by floating the source to eliminate current flow.

4.2.3.4 Read

An element is read by grounding the source, raising the control gate, and sensing current flow as the drain is raised to approximately 1 V. The element that is physically adjacent to the element being read and that shares the same drain and control gate can source current to the bit-line (metal line connecting common drains). The adjacent element can also sink current from the bit-line. These currents are avoided by floating or driving the source of the adjacent element such that the source approaches the potential of the element's drain.

4.2.3.5 Disturbs

The X-cell is, of course, subject to all of the same disturbs as the T-Cell. In addition, dynamic disturbs that arise as a result of capacitive charging and discharging of the bit-lines due to the virtual ground configuration must be considered.

4.2.4 Staggered virtual ground (SVG[®]) array UV-EPROM cell⁸

4.2.4.1 Description

The SVG array is discussed in Ali, et al. [B2]. A virtual ground array is an attractive area-saving alternative to the common ground array. It also reduces the number of contacts in the memory array. The cell in the SVG array is the self-aligned split-gate cell. The cell cross-section is shown in Figure 21. It is composed of an MOS transistor in series with a floating gate transistor that is merged into a single composite device (Eitan [B19]). The second layer of polysilicon acts as the control gate and directly covers part of the channel area. The n+ bit-line on the drain is formed by an ion-implant step that is aligned to the floating gate. Thus, the channel length of the floating gate transistor is mask alignment insensitive. The channel length of the series MOS transistor is determined by the distance between the edge of the floating gate and the n+ bit-line, and varies with misalignment of bit-line mask with respect to the poly 1 (floating gate) layer.

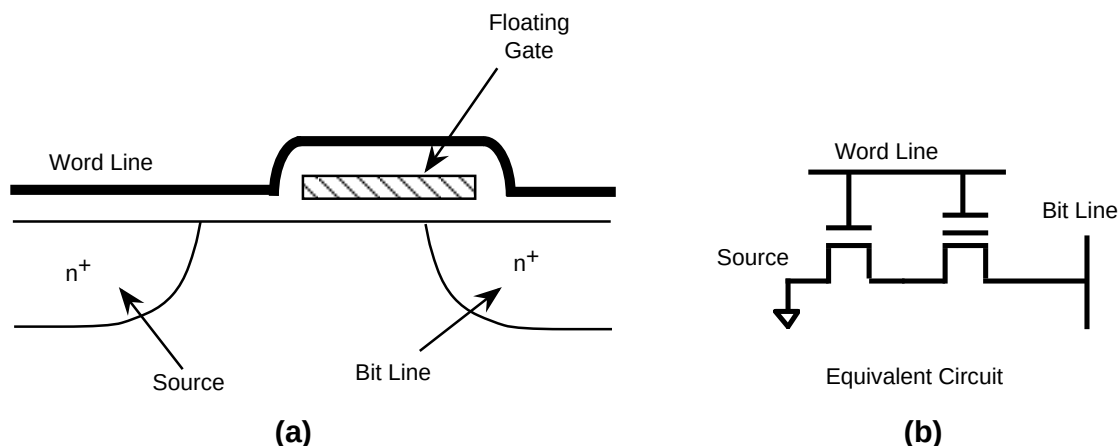


Figure 21—(a) Cross-section and (b) equivalent circuit of the self-aligned split-gate cell

⁸SVG is a registered trademark of WSI.